

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2004, 05:42:58 ; Search time 59 Seconds

(without alignments)
6850.753 Million cell updates/sec

Title: US-09-889-874A-23

Perfect score: 8879

Sequence: 1 VIKELKLFRRITMSDNEF.....PRKILGRTEKTKPKTERP 1673

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8879	100.0	1673	3	Aay95707 Cosmid cH
2	1638	18.4	1584	2	Aay33727 Photorhab
3	1583	17.8	1787	6	Abm67433 Photorhab
4	1569	17.7	1590	6	Abm67283 Photorhab
5	432	4.9	2334	5	Abg31849 Human kin
6	386	4.3	2234	6	Abu18641 Protein e
7	365	4.1	1400	7	Adc01365 Enterohae
8	353.5	4.0	843	6	Abu48290 Protein e
9	350	3.9	1515	6	Abu19676 Protein e
10	342	3.9	1397	6	Abu15135 Protein e
11	341.5	3.8	1404	7	Adc00960 Enterohae
12	341	3.8	1394	7	Adc01426 Enterohae
13	340.5	3.8	1411	6	Abu15136 Protein e
14	339	3.8	1426	3	Abu15983 E. coli p
15	339	3.8	1426	6	Abu14693 Protein e
16	331.5	3.7	1377	4	Aau34791 E. coli c
17	331.5	3.7	1377	4	Aag98997 E. coli g
18	331.5	3.7	1377	6	Abu14807 Protein e
19	328	3.7	1572	6	Abu41491 Protein e
20	324.5	3.7	1429	6	Abm69806 Photorhab
21	322	3.6	1385	6	Abu40318 Protein e
22	317.5	3.6	1584	6	Abm66973 Photorhab
23	317.5	3.6	1585	6	Abu16634 Protein e
24	314	3.5	1395	6	Abu22662 Protein e
25	313	3.5	1627	6	Ada35317 Acinetoba

26	307.5	3.5	932	6	ABU15693	ABU15693 Protein e
27	306.5	3.5	1530	6	ABU40251	ABU40251 Protein e
28	298.5	3.4	1531	6	ABU21384	ABU21384 Protein e
29	298	3.4	885	4	AAU52772	AAU52772 Propionib
30	298	3.4	885	6	ABM49291	ABM49291 Propionib
31	298	3.4	920	6	ABM65127	ABM65127 Propionib
32	297	3.3	1512	6	ABU50462	ABU50462 Protein e
33	291.5	3.3	1253	6	ABU40990	ABU40990 Protein e
34	289.5	3.3	1364	6	ABU47385	ABU47385 Protein e
35	285.5	3.2	1596	6	ADA34255	ADA34255 Acinetoba
36	285	3.2	2613	5	ABP53589	ABP53589 Human NOV
37	285	3.2	2628	5	ABP53588	ABP53588 Human NOV
38	285	3.2	2721	5	ABP53587	ABP53587 Human NOV
39	285	3.2	2725	5	ABP53586	ABP53586 Human NOV
40	280	3.2	1317	4	AAU33622	AAU33622 pseudomon
41	280	3.2	1317	6	ABU15571	ABU15571 Protein e
42	279.5	3.1	944	6	ABU19882	ABU19882 Protein e
43	278	3.1	985	6	ABU15137	ABU15137 Protein e
44	276.5	3.1	1438	6	ABU50436	ABU50436 Protein e
45	273.5	3.1	1688	6	ABR58344	ABR58344 XM_047995

ALIGNMENTS

RESULT 1

AA95707

ID AAY95707 standard; protein; 1673 AA.

XX AAY95707;

DT 25-OCT-2000 (first entry)

XX Cosmid CHRIM5 encoded protein P14-2f.

XX Cosmid CHRIM5; nematocite; nematode; biological control agent;

XX transgenic plant; helminthiasis; P14-2f.

XX Xenorhabdus bovienii.

XX WC2000042855-A1.

XX 27-JUL-2000.

XX 24-JAN-2000; 2000WO-GB000219.

XX 22-JAN-1999; 99GB-00001499.

XX (HORT-) HORTICULTURE RES INT.

XX Morgan JAW, Jarrett P, Ellis D, Ousley MA;

XX WPI; 2000-499157/44.

XX N-PSDB; AAA50029.

XX Novel composition used to control parasitic nematodes, especially in a plants such as maize, cotton, soya, and rice, comprises a bacterium which is a symbiont of an entomopathogenic nematode.

XX Example 6; Page 42-43; 74pp; English.

XX The present sequence is that of protein P14-2f encoded by an open reading frame identified in cosmid CHRIM5 (see AAA50029). CHRIM5 was obtained by ligating Xenorhabdus bovienii strain 173 (NCIMB 40986) Sau3A-digested DNA fragments into the BamHI site of the Strategene cosmid vector Supercoiled, packaging into Escherichia coli XL Blue 1, and screening for nematocidal activity against Caenorhabditis elegans. Analysis of the DNA indicated a number of open reading frames for which the corresponding protein sequences were determined (see AAY95685-Y95735). Nematodes can be controlled through the use of bacteria associated symbiotically with an entomopathogenic nematode. Such bacteria include Xenorhabdus and Photobacterium spp. such as X. bovienii strain 173. The symbiont bacteria, an engineered bacterium, or a nematocidal protein obtained from such

CC bacteria, particularly P13-1f (see AAY95706) or P14-2f can be used to
 CC control helminthiasis in a human or domesticated animal or for the
 CC control of plant pathogen nematodes. Also claimed are vectors for
 CC expressing nematocidal proteins in host cells, and transgenic plants
 XX
 SQ Sequence 1673 AA;

Query Match	100.0%;	Score 8879;	DB 3;	Length 1673;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1673;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	VYIKFLPFRITWSDNNNEFFTCANNFTSAVSGGVDPRGLYNIQITILGHVGNGLGPT	60	
DB	1	VYIKFLPFRITWSDNNNEFFTCANNFTSAVSGGVDPRGLYNIQITILGHVGNGLGPT	60	
QY	61	LPTLSYSPLNKTDIGFIGNFGLSVYDRKNSLSLSSTGENYKVIETDKTKVLQKKLD	120	
DB	61	LPTLSYSPLNKTDIGFIGNFGLSVYDRKNSLSLSSTGENYKVIETDKTKVLQKKLD	120	
QY	121	NLRFEXDLKENCYRIITHKSGDIEVLTFGNNAFDLKVPKLLNPAGHAIYIDMFEATOP	180	
DB	121	NLRFEXDLKENCYRIITHKSGDIEVLTFGNNAFDLKVPKLLNPAGHAIYIDMFEATOP	180	
QY	181	RLNRIYDDLGDHIDPLINLEYQSLIKTILTLFPQKQGYRTELRLFLNQLNSIHNSLGN	240	
DB	181	RLNRIYDDLGDHIDPLINLEYQSLIKTILTLFPQKQGYRTELRLFLNQLNSIHNSLGN	240	
QY	241	ENPLTWSFGYTPIGKXGILGOWITMTAPGGLKETVYNSNNQGHFFQSANLPVLPVYT	300	
DB	241	ENPLTWSFGYTPIGKXGILGOWITMTAPGGLKETVYNSNNQGHFFQSANLPVLPVYT	300	
QY	301	LMQVPGAGOPAIQAEYSYTSNHYVGGSGNGIWNKLNLYGLMTEYNYGSTESEYKXK	360	
DB	301	LMQVPGAGOPAIQAEYSYTSNHYVGGSGNGIWNKLNLYGLMTEYNYGSTESEYKXK	360	
QY	361	EGHDQIVIRIERTYNNYHLLTSECKQNGVIOITETAYYAIIGHNFDSPQFQPKTKTE	420	
DB	361	EGHDQIVIRIERTYNNYHLLTSECKQNGVIOITETAYYAIIGHNFDSPQFQPKTKTE	420	
QY	421	TWRSADNSVSEITETTFDESNGNPLTKVIKDKTKQKIISPSHWHYPPAGEVNCNCPPEP	480	
DB	421	TWRSADNSVSEITETTFDESNGNPLTKVIKDKTKQKIISPSHWHYPPAGEVNCNCPPEP	480	
QY	481	YGTRFVKKIIQTPYDSEFKDDPEKFIQVRYSLIGSQSHVTLKIBERHVSATQLNLSLTF	540	
DB	481	YGTRFVKKIIQTPYDSEFKDDPEKFIQVRYSLIGSQSHVTLKIBERHVSATQLNLSLTF	540	
QY	541	QYNTDKSELGRLLKQTECKGNGKTYSVVHFTYTKODDTLQOQSHSITTHDNFTIHSQ	600	
DB	541	QYNTDKSELGRLLKQTECKGNGKTYSVVHFTYTKODDTLQOQSHSITTHDNFTIHSQ	600	
QY	601	VRBRYTGRLPDSDTDKDIIVTQMSYDKLGRLLRTNLNSGTPYANTLTYDELNNLQDDNRP	660	
DB	601	VRBRYTGRLPDSDTDKDIIVTQMSYDKLGRLLRTNLNSGTPYANTLTYDELNNLQDDNRP	660	
QY	661	PFVITTTDVGNGQLRNEFDGAGRHSVQCLKDSGDKGFYTHIQOYDEQGRHHTSTYSYD	720	
DB	661	PFVITTTDVGNGQLRNEFDGAGRHSVQCLKDSGDKGFYTHIQOYDEQGRHHTSTYSYD	720	
QY	721	LITNGRQOTDPDKVHLGMSKSYDNWQGIANTHWSYGVSEKIVDPIITLTKLOLSNSNV	780	
DB	721	LITNGRQOTDPDKVHLGMSKSYDNWQGIANTHWSYGVSEKIVDPIITLTKLOLSNSNV	780	
QY	781	QTKQEVITYPPSQOPIQITLPEAGHLQSCHTLTRDGDWRVREKETAIGCQTIYQYDNYN	840	
DB	781	QTKQEVITYPPSQOPIQITLPEAGHLQSCHTLTRDGDWRVREKETAIGCQTIYQYDNYN	840	
QY	841	RVLIQITLPDGTIVNRKVPFSDTLTDIRVNGISLIGQOTPDGLSRLTQSGDGRVWAYT	900	
DB	841	RVLIQITLPDGTIVNRKVPFSDTLTDIRVNGISLIGQOTPDGLSRLTQSGDGRVWAYT	900	
QY	901	YSAGNDQCPSTVITPDGQFIHYQOPELDDAVLQVASNEITQOFSYNPVTGALLKAVAG	960	

DB	901	YSAGNDQCPSTVITPDGQFIHYQOPELDDAVLQVASNEITQOFSYNPVTGALLKAVAG	960	
QY	961	QSLTPIYPSGRKLMENINDMKMSYLTWLRGLNGYTDLTGTTQIKISRUTHGRVTOIKD	1020	
DB	961	QSLTPIYPSGRKLMENINDMKMSYLTWLRGLNGYTDLTGTTQIKISRUTHGRVTOIKD	1020	
QY	1021	SSIKITLNVDDLNRIHIGSQVTDLATGCHMLTTTTFVDFGLNRIEGRKLCSSGHTLIDQOSW	1080	
DB	1021	SSIKITLNVDDLNRIHIGSQVTDLATGCHMLTTTTFVDFGLNRIEGRKLCSSGHTLIDQOSW	1080	
QY	1081	LKTQOLANRIVKLNGVQRTQYSYDSNRNLNQVKDGAECPTDKYGHISIVTQNTFTYDIY	1140	
DB	1081	LKTQOLANRIVKLNGVQRTQYSYDSNRNLNQVKDGAECPTDKYGHISIVTQNTFTYDIY	1140	
QY	1141	GNITACHITFADGTEDHATFKFANPTDPCQLTEVHTHPPMDPNIRLKYDAGRVINITD	1200	
DB	1141	GNITACHITFADGTEDHATFKFANPTDPCQLTEVHTHPPMDPNIRLKYDAGRVINITD	1200	
QY	1201	NHGNTEFTYDPLGRQLQNGQSVYGYDPLNRLVSKQTDLTDCELYRYRETMVNEVRNGEM	1260	
DB	1201	NHGNTEFTYDPLGRQLQNGQSVYGYDPLNRLVSKQTDLTDCELYRYRETMVNEVRNGEM	1260	
QY	1261	IRLLRTGETIIAQQRASKVLLTGTDSQOSVILTSQKQNSOEAYSAYGKHSTANDASIL	1320	
DB	1261	IRLLRTGETIIAQQRASKVLLTGTDSQOSVILTSQKQNSOEAYSAYGKHSTANDASIL	1320	
QY	1321	GYNGERADPVSGVTHLNGYRSYDPTLMRFHTPDSLSPPFGAGGINPYSCIGDPINRSDP	1380	
DB	1321	GYNGERADPVSGVTHLNGYRSYDPTLMRFHTPDSLSPPFGAGGINPYSCIGDPINRSDP	1380	
QY	1381	SGHLSQWATGIGMGIAGLLTIATGMAIAAAGIAAAASTTTALAFALSVTSBIT	1440	
DB	1381	SGHLSQWATGIGMGIAGLLTIATGMAIAAAGIAAAASTTTALAFALSVTSBIT	1440	
QY	1441	SIVSGALEDASPASSIILGWVSMGMAAGLAESAIGKGTKLATHLGAFAEDGENALLKST	1500	
DB	1441	SIVSGALEDASPASSIILGWVSMGMAAGLAESAIGKGTKLATHLGAFAEDGENALLKST	1500	
QY	1501	SSSRIRKGVTRSLDREIVRNEEQVTKDHSRGYTDNFMKGEOAILVHGDKGFLYHTE	1560	
DB	1501	SSSRIRKGVTRSLDREIVRNEEQVTKDHSRGYTDNFMKGEOAILVHGDKGFLYHTE	1560	
QY	1561	GNKNGKGPYRHTPEQLVDYLKDNINIVDLTQGGDKPVLHLLSCYKSSGAADKQAKYINR	1620	
DB	1561	GNKNGKGPYRHTPEQLVDYLKDNINIVDLTQGGDKPVLHLLSCYKSSGAADKQAKYINR	1620	
QY	1621	PVIAYSNKPTISQGLARIEKDFLKSTYHSYDPRKILGRTEKTVKPKTRP	1673	
DB	1621	PVIAYSNKPTISQGLARIEKDFLKSTYHSYDPRKILGRTEKTVKPKTRP	1673	

RESULT 2

AAV33727	
ID	AAV33727 standard; protein; 1584 AA.
XX	AAV33727;
AC	AAV33727;
XX	
XX	09-NOV-1999 (first entry)
DT	
XX	
DE	Photobacterium luminescens 176 kD insecticidal toxin.
XX	
KW	Symbiotic bacterium; nematode; insect; larva; toxin; insecticide.
XX	
OS	Photobacterium luminescens.
XX	
FN	W09942589-A2.
XX	
PD	26-AUG-1999.
XX	
PF	18-FEB-1999; 99WO-EP001015.
XX	
XX	20-FEB-1998; 98US-00027080.
PR	20-JAN-1999; 99US-0116439F.

Db 1464 SQAVSAGVIGVPLEF-----GEVASRSSR 1488

RESULT 3
ABM67433
ID ABM67433 standard; protein; 1787 AA.
XX AC ABM67433;
XX DE 20-NOV-2003 (first entry)
XX DT Photorhabdus luminescens protein sequence #530.
XX DE
XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough.
XX OS Photorhabdus luminescens.
XX PN WO200294867-A2.
XX PD 28-NOV-2002.
XX PF 07-FEB-2002; 2002WO-IB003040.
XX PR 07-FEB-2001; 2001FR-00001659.
XX PA (INSP) INST PASTEUR.
XX PA (CNRS) CNRS CENT NAT RECH SCI.
XX PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;
PI WPI; 2003-148459/14.
XX DR
XX PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX PS Claim 2; SEQ ID NO 530; 1205pp; French.
XX CC The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterial agents useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX SQ Sequence 1787 AA;

Query Match 17.8%; Score 1583; DB 6; Length 1787;
Best Local Similarity 28.2%; Pred. No. 2.4e-101;
Matches 511; Conservative 280; Mismatches 727; Indels 294; Gaps 65;

QY 22 TQANNFTSAVSGGVDPRGTGLYNIQITLGHVGNGLGPTLPLTILSYSPLNKTDIGFGIGF 81
DB 30 SQAINFTGAMWGVDPRTGLTQIPITQIGGDLWGPDLAILTYNPLTYLNTGYGLGF 89

QY 82 NFGLSVYDRKNSLLSLSTGENYKVIETDKTV-----KLOQKKLDNLRFEKOLKENCYRI 135
DB 90 SDNFTRYDTOTQVLTATGSEIYHREKANEVVDQAWTFHAKPAHFKVKE--KDAFWV 147
QY 136 IHKSGDIEVLTFGNNAFADLVKPKKLLNPAGHAIYIDWNFEATQP-----RLNRYDDLDG 191
DB 148 LYKTSREKLTQL--DRANPVAIVSEIYAPSGHKLCKWNSFVNHNYYNWLMEVCDAME- 205
QY 192 HDIPLNLLEYQGLIKITILTFPGQKEGYRTELAFNLQNLNSIHNFSLGNENPLTWSEGYT 251
DB 206 ---TLKADLATEKLEFTVWPGSPESYTVTLNNTDLLQTVISAS-----ELTWLEYE 257
QY 252 PIGKNGILGOWITSMTAGGLKETVYNNNNQGHFP----- 288
DB 258 TEGAHKNI--LTKVTTPSGLIEKVVY--HETGHTLTPKCIQYFNAMWPGIIRDPKS 312
QY 289 -QSANLPVLPVTLMKQVPGAGOPAIQAEYSYTSNHYVGGSGNGIWN---NKLNLNLYGL 343
DB 313 STTATVEHFPYVTHHIIAGSSGSPDQVIRVVFSPENFLGOKNWKMDPIPLPODNAYLA 372
QY 344 MTEYNVGSSTESRYKKEGHDOIVRIERTYNNYHLLTSECK-----QONGYIOTTETAY 397
DB 373 NSEYKYSTEVREYNEKR-----YCHREYNKPHLLVSETEVETVTPSRPQKLKTIKY 427
QY 398 YALIIGHNFD--SOPSOFLPKTKTETWRSADNS-----YRSIETETTFDESIGNPLTKV 450
DB 428 YADVGSFDDNKOPQFLMPTVETIWHNPESASTQRKETTOWEYNAOCLNLSMTLP 487
QY 451 DKTKQKIISPSTHWEYYPAGEV---DNCPPPYGFTREVKKI-IQTPYD-SEFKDDPEK 505
DB 488 DNTTKT-----TYAPDGEETDTHCPAEPNGFERIKEIAVEAPSLAKTKITILR 540
QY 506 FIQYRSLIGSQS-----HVTLEIRHYS-----ATQL-----LNSTLFOYNTD 545
DB 541 KVTINYKSDTSPPKNOASVVKNSVLPSTHYSRRCCADHLEKVKINTVSFVENTQ 600
QY 546 KSELGRLLKQTECKGENG-KTVSVVHKFTYTKQDDTLQQSHSITTHDNF-----TIHRS 599
DB 601 NAFNLRVBOURNSYLLFENGNGQTRSYTEDYSWSENKNGASCICKKTGSGKGTTPVSH 660
QY 600 QVRSRYTGBLFSDDTKDIVTOMSYDKLGRLLTRNLNGSTPYANTLTVDYELNNLQDN 659
DB 661 QVWSRSTGBLIFOKAQDNNTVFQYDTTGLHLSSTINADTAYEKTVKYAWSYNSKR---- 716
QY 660 PPFVITTTDVNGQLRNEFDGAGRHVSCLKSDGDKFYTHITQQYDEQGRHHTSTYSD 719
DB 717 --VTVTQDIHENRYITEMDGLGRPLKXSYPAGYQGGQFOMERYQYNPLGOLQOASCD 774
QY 720 YLTNGRQOTDPKVLHLSKSYDNWNGQIANTHWSYGVSEKITVDITLTATKQLQSN 779
DB 775 HML-FENKSEKCCSMTVSLEYDDWGHNSRLYSDDGTSVQNTFDPKMTSEHQRVSNDES 833
QY 780 VOTGKEVTTYTPSQOPIQITLFDAGHLQSCHTLTRDCMDVRKETDAIGCTIVQYDNY 839
DB 834 QSSGIIRTYNPGFQITATERLTTSIQGCHWYLRDLGRLL-VSINANGNTLLAYDAF 892
QY 840 NRVIOITIPDGHVNRKYA-PESTDTLITDIRVN--GISLQOQTFDGLSLRUTQSDGGRV 896
DB 893 DRVIRQTFADGTTISMAVENGVSVMSATPLGINTQPVILGTQILDGLGRVIDMESGGRK 952
QY 897 WAYTYSAGNDQCPSTVITPDQ-----FIHYQVQPELDDAVLQVAS--NEITQQESYN 947
DB 953 IKLDYE-GASVPVDVTYVKGTSQDKDVLHYEYEPKLNNAVTKITAGQDDVDQOTFKYD 1011
QY 948 PVTGALLKAVAF-----GQSLTPIYPSGRKME-----NINDMKMSYLTWLR-G 992
DB 1012 PKTGLLTAIEYKKNKNDKEMISSKLIIFYTLSSGLASEKLVSNKNKNRMYHFKTCSFS 1071
QY 993 LENGVTDLTGT-----IQKISRDTGRTVQIKDSSIKITLNYDDLNHIGSQVTDL-- 1043
DB 1072 FANRPTAKSTINGPRYIINKYSYDNQGRLETVERNDIEINLKYKLSRHVVCQSTYFLD 1131
QY 1044 ATGHLMTTTFVDFDGLNREIGRK-----LCDSSGHT-LDIOQSWLKTQQLANRIVKNGV 1096

Db 1132 TINNKITSLDFOGRETETKIRRHNGLKQNKGITDIHSQTFNEQDKITNKL-LHGT 1190
Qy 1097 LQ-RTEQSYSDRNRLNQKCDGAECPDCKYGHVLTQ--NFTYDVIYGNITACHTTTADG 1153
Db 1191 KQISKEVYTHKRGLETYT-----MELVGEREITQCSYKDYHGLNITQ-HSITTEG 1243
Qy 1154 TEDHATKFNAPT--DPCQLTEVHH-THPDMPDNIRLYKDKAGRVINITD-NHGNTENFT 1209
Db 1244 XTITSTYTGNGNIQDPCQLIDVSTCTGNNSSSLRFTYNGQALVCENDENNTKIRWT 1303
Qy 1210 YDTLRLQNGCGSV-----YCYPLRLV---SQKDET-LDCELYYRETMVYN----- 1253
Db 1304 YDSLGRIDTVDALFKVETRYFLDATRLIIRKSEKNGTPVHHDLSCYNSLVHNDVYFG 1363
Qy 1254 EVRNGEMLR-----LLRGETIIAQ--QRASKVLLITGDSQQSVILTSKQNLQSOE 1302
Db 1364 EKRONAADRKYKNGVGGICLGFQSFCHQTPTSVASRYTETATDCKGVIATFQGEDVQHI 1423
Qy 1303 AYSAYGKHSTANDASILG-----YNGERADPVSGVTHLNGYRSYDPTLM 1348
Db 1424 AYSFPGW---VTTEQAMVTAGQPPHNTAIEBPRFNGEQMDTASASYLLNGYRAYRPDL 1480
Qy 1349 RFHPDLSLPRGAGINPVSYCLGDPINRSPSGHLSWQAMTGIGMGIAGLLLTIATGCM 1408
Db 1481 RFTAPDSWSPGAGINAYACGGDPVNLNDPSGHSIGWGANITGGIGLLAPFTYVG 1540
Qy 1409 AIAAGGIAAIASTTTALAFGALSVTSDITSIVSGALEDASPASISILGWGMGAA 1468
Db 1541 SLELGLGVNAARGLT-----ALDAASGVTAASGALENKNPETSRRLGWMBSLGLG-- 1590
Qy 1469 GLAESAIGKTYKLA-----THLGA-----AEDGENALLKKTSE--- 1502
Db 1591 --LPSMWITGGVSLAOWNVRLNTPRTPYHPTSLGEVNLNKRKSDWVNAARSLNSENW 1648
Qy 1503 -----SSRIKWG---VTRSLDREIVRNEEQVINDHSGRYTDNFMKGEOAI-----L 1547
Db 1649 HSEVLNGRTIWGSDTKRGLD---IKYPLEQISGRPSNG--DIVLLSGSHGVQNGDWL 1703
Qy 1548 VHGDKDGFLYH-----TEGNKHGKGPYTRHTPEQLVDYLDKNNIVDLTQGGKPV 1598
Db 1704 INGSRRGSLHFPFPGKDMTYGSGWKGRTHVRN-----LATWSEIDFGTLNNGNSHI 1758
Qy 1599 HLLSCYKSSGA 1610
Db 1759 ILCYCYGRNDQA 1770

RESULT 4
ABM67283
ID ABM67283 standard; protein; 1590 AA.
XX ABM67283;
AC ABM67283;
XX 20-NOV-2003 (first entry)
DT
XX Photorhabdus luminescens protein sequence #380.
DE
XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough.
XX
OS Photorhabdus luminescens.
XX WO200294867-A2.
XX
XX 28-NOV-2002.
PD
XX 07-FEB-2002; 2002WO-IB003040.
PF
XX
XX 07-FEB-2001; 2001FR-00001659.
PR
XX

PA (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
XX Buchrieser C;
XX WPI; 2003-148459/14.
XX
PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
PS Claim 2; SEQ ID NO 380; 1205pp; French.
XX
XX The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX
SQ Sequence 1590 AA;
Query Match 17.7%; Score 1569; DB 6; Length 1590;
Best Local Similarity 30.8%; Pred. No. 1.9e-100;
Matches 477; Conservative 236; Mismatches 636; Indels 200; Gaps 57;
Qy 63 LTUSYGPLNKTDIGFQIGFNGFGLSVVDRKNLSLSLTGENYK---VIETDKTVLKQOKKL 119
Db 13 LNLFSPLATLNNFGFGIGWRFSLTMLDVKTLTFRSNGEQFKCKPLPPNNNDISFKDKL 72
Qy 120 DNLRFKDKLKNYRIIHKSGDIEVLTFGNNAFDLKVPKLLNPACHAIYIDWNEATQ 179
Db 73 KDLRVTK-LDSNFFVYVNGIETLKRGSS--DIAKTVALEFPDG----- 116
Qy 180 PRNLRIYDDLGDHDIPLNLLEYQGLIKTILTL-FPGQKEGYTELRL-NRQLNSIHNS 237
Db 117 ---EVFDLTYNRFRFALSEIKYRMTGKTYLKLNSGNN---CTSVEYPPDNNISAKIAFD 169
Qy 238 LGNENPLTWSFGYTPIGK-----NGILGQWITSMTAPGLKETVYNSNNQ 284
Db 170 YRNDYDITVTVPYDASGPDIDSAFKNYQTLKGIF-FVISAFTPTGYVELSVYKEN--G 226
Qy 285 HHFPQSANLPVLPVYTLMKQVPGAGQPAIQAEYSYTS-HNYVGGSGNGI-WNNKLDNLYG 342
Db 227 HKYTDTEISIPYAAALTIQ---PCNGQPAISKSYEYSSVHNFGLYSSGRTSFDSSQDNLYL 283
Qy 343 LMEYNYGSTESRRYKDKGHDQIVRIERTYNNYHLLTSECKQONGYIQTETAYVAILG 402
Db 284 VTGKITYSSIE-----RYLNGQNVISVTERVDFKHLMTKEAKTDQNKRIITEITYNEDPS 339
Qy 403 HNFDSQPSQFQLEK-----TKTETWRSADNSYRSEITETTFDESQNPITKVIKDKTKQ 456
Db 340 KSPSEQENLQPSHVLTRYTDLOT-----NTSREESVNIKSDDWGNTLL-ITETSGIQK 393
Qy 457 IISPSHWEYYPAGEVNDNCPPEPYGTRFVKKIIQTPYDSEFPDKDPEKEIQRYSLIGS 516
Db 394 -----EYVYVPVNGEGNCPADFLGFSRFLKSTQKSPDAAQSVANRVTSYQKLT 447

QY 517 QSHVTLKIEERHYSATQLLNSTL---FOY---NTDKSELGRLLKOTECTKGNGKTSYVW 570
 Db 448 FTGAYVK---EYYSKASETIDSKIVRTFNVSFTKSH---GSLAKITSWNNO-----QTV 499
 QY 571 HKFTYTKQDDTLQOQSHSIITHDNFTIHRSOVSRVYGRFLFSDDTKDIIVTQMSYDKLGLRL 630
 Db 500 TTFKYEYSDSEMTNSTVTFGFGTHMESKNVTSIVTHRQLRKVDVNVHVTIDQSYDLSGRI 559
 QY 631 LPTLNSGTFYANTLYDYELANLQDDNRPPEVITTTDVANGNQLNEBPDGAGRHVSQCLK 690
 Db 560 IGIIDDPGTTKEIKSIIYYPCGGENDFWP---VMEIDSQGIIRKTHYDGMGRICISIEQ 618
 QY 691 DSDG-----DGFTYTHITQYDEQGRHHTSTYSVLTN-----GRQOTDPDKVHLISM 737
 Db 619 DDDGVWGTSGIYQGYRKLARQYDVLGQVKEISNDMLWDLNPLRLTLP---LVITK 676
 QY 738 SKSYDNWQOLANTHSYGVSEKLTVDPIITLPAK-----QLOSNNVOTGKEVITTP 791
 Db 677 TYQYDGNRNYSTESDGRIELEIHDPITRTITQGVKGLGMLNIQNNF-----725
 QY 792 SQPIQITLFDGAGHLSCHTLTRDQWDRVRKETDAIGQCTIYOYDYNRVIOITLPDGT 851
 Db 726 -BQASIKVYPDGAISYRTYRVDGFGTVTETDAEGYATQIEYDLDFRIVKXLTLPRT 784
 QY 852 IYVRKYAPSTDTLITDIRVNGISIGQOTFDGLSRLTOSQDGRVWAVYTAGNDQCPST 911
 Db 785 ILESAYASFSEELISALNVNTOGLSIVYDGLGRVTRDTVGRKTEYLYGSQDK-PIQ 843
 QY 912 VITP---DGFYHVOQPELDDAVLOVANEIITQFSSYNPVTGALLKAVAEGOSITFI--YY 968
 Db 844 SVTPAHKKNIDLY--ALGWSMFTTETSQNSFSYQKTCALLSA--TEGVSGNSYF 900
 QY 969 PSGLKLMENIN-DMKMS---YLWTLRGLNGYDITLGTIQIKSRDTHGRVTOIKDSSIK 1024
 Db 901 PSGLVCHESFSDNKPISGSDYRYTMSGLIQSHKSFADHVSVDAGRLVKTEQSSQY 960
 QY 1025 TTLNVDDLNRHGSQVDTLATCHMLTTVEPDGLNREIGRKLCDSSGHTLIDQOSWLTKQ 1084
 Db 961 ATFEDYDNGRLTITTTKDTTSLSQATKIEYDFVDFREIKRSLISDFSQV--ITLSYTKN 1019
 QY 1085 QLANRIVKLVGLQRTQYSDSRNLNQYKDGACEPTDKYHSIVTQNFYDIYGNIT 1144
 Db 1020 QISQRTSIDGVVMKNERYOYDSNQRLSQYQCEGQSVDHTGRVLSQIIVHYDQWGNIK 1079
 QY 1145 ACHTFADCTEDHAFKANPTDPCQTEVHTHDPMDPNILKXDKAGRVINIT-DHNG 1203
 Db 1080 RLNDTYRDKET-VDYHFSQ-ADPTQLIRI-----TSDKQQLSELYDANG---NLTRDEKG 1130
 QY 1204 NTENFTYDITLGLR---QNGQGSV---YGYDPLNRLVSQ--KTDTLDCELYRYETMLVNEV 1255
 Db 1131 QT--LIYDQNNELVQVQSKGNLVQCYQYDALNKLTAQVLANGTVNRQ--VYASGNVANVQ 1187
 QY 1256 RNMGMIRLL-----RTGTIIAQORASKVLTGTDSQOSVILTSDKNLQOBA 1303
 Db 1188 LGDETITLMSDKQRLGHQSTKNQSGSVYQY-----GTDHNSVTIASQNELMALS 1239
 QY 1304 YSAYGKHSTANDASILGVNGERADPVSGVTHLNGYRSYDPTLMRPHTPDSLPFGAGG 1363
 Db 1240 YTPYGRSLI---SSLPLNGAQVDPVTGTFELNGYEVFVFLMRPHFSPDSWSFFGRGG 1296
 QY 1364 INPYSYCLGDPINRSDPSGHSLSQAWTIGIGIAGLLTIAT--GWAIA-----AAGGIA 1417
 Db 1297 VNPYTYCQDPINRIDNLGHLSAGGILGIVLGAIGIIVGVVSLGAGAISAAGLAAAGAL 1356
 QY 1418 AAIASSTTALAFAGALSVTSDITTSVSGALDEASPASSIILGWSMGMAAGLAESAII-- 1475
 Db 1357 GAIASTSAFATVATVIGLAASIGIASAALSEKDPKTACILNWIISTGLCVSPGISAITP 1416
 QY 1476 -----KGGTKLATH--LGAF-AEDGENALLKSTSESSRIKMGVTRS 1513
 Db 1417 TSLIKSARSGSVASTSVIGSVPIEFGEIA-----SRSSR-RWDIALS 1459

RESULT 5
 ABG31849
 ID ABG31849 standard; protein; 2334 AA.
 XX
 AC ABG31849;
 DT 05-NOV-2002 (first entry)
 XX
 DE Human kinase, MEK1.
 XX
 KW Extracellular signal regulated kinase; hyperalgesia; surgery;
 KW opioid withdrawal; pain sensitisation; analgesic; chronic pain; BRK;
 KW MEK1; human; enzyme.
 XX
 OS Homo sapiens.
 XX
 EN W0200258687-A2.
 XX
 PD 01-AUG-2002.
 XX
 PF 25-JAN-2002; 2002WO-US002128.
 XX
 PR 25-JAN-2001; 2001US-0264336P.
 XX
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Gutstein HB;
 XX
 DR WPI; 2002-608414/65.
 DR
 DR N-PSDB; ABK30804.
 XX
 PT Reducing or reversing tolerance, physical dependence, hyperalgesia,
 PT withdrawal symptoms, or pain sensitization in patients on analgesics for
 PT chronic pain, comprises inhibition of the extracellular signal-regulated
 PT kinase (ERK).
 XX
 PS Disclosure; Page 156-161; 163pp; English.
 XX
 CC The invention relates to a method of reducing or reversing tolerance,
 CC reducing the risk of physical dependence or hyperalgesia, reducing the
 CC symptoms of opioid withdrawal or inhibiting pain sensitisation in a
 CC patient taking analgesics. The method comprises administering an
 CC analgesic and an extracellular signal-regulated kinase (ERK) inhibitor
 CC comprised in a formulation to reduce or reverse tolerance, risk of
 CC physical dependence, hyperalgesia, symptoms of opioid withdrawal, or
 CC inhibiting pain sensitisation in patients taking analgesics for chronic
 CC pain or those undergoing surgery. The present sequence represents the
 CC amino acid sequence of human MEK1 (not defined)
 XX
 SQ Sequence 2334 AA;
 Query Match 4.9%; Score 432; DB 5; Length 2334;
 Best Local Similarity 20.4%; Pred. No. 3.3e-20;
 Matches 358; Conservative 226; Mismatches 613; Indels 560; Gaps 80;
 QY 44 IQITLGHIVG-NGNL-----GPTLPLTSLYSPLNKTIDIGFIGNFGLSVYDR 90
 Db 896 IDIPSQLNGATGATVNVNEEDLSIDGRPGGLGSLRTYNSLDSSDHLFGQGW-----YADA 950
 QY 91 KNSLLSLSTGENYKVIETDKTVLQOKKLDN-----LRFKDLKENCYRIHKSGDIEVL 145
 Db 951 ETSVISTDQAMV--IDEDATTHRTFKADGTGYQPTGYVLELTADQFILKTKD---- 1004
 QY 146 TGFNNNAFLKVPKLL-----NPAGHAIYIDWN-----FEATQPLNRIRYDDLDGH-- 192
 Db 1005 ---QTNAYNKKGKLVQVVDGHNNAVTVYNDKNQLTALTDASGRKLTFTYDE--NGHWT 1060
 QY 193 -----DIPLLNLEY--QGLI-----KTIITLFPG 214
 Db 1061 SITGPKNKVTSYENDLLKKVYTDGTDTGVTSDYDSEGLRVKQYGSANSTKAPVFTEY-- 1118
 QY 215 QKEGYSTELRFLNRLNLSIHNS-----LGNENPLTWSFGYTPIGKNGILGQWITSMT 267

Db 1119 QYSHRLEKAINAKETVYVSYDADKKTLLMTPQNGRKVOYGYNEAGNP-----IQVID 1172
QY 268 APGLKETVN--YNNNN-----QGHFPQSANLPVLPVYTLMKQVPGAG 309
Db 1173 DAELGLKITTNTKYEGNNWVEDVPNDVGTGKATESYQYDKDGN-----VTSVKDAYGT- 1225
QY 310 QPALQAEYSYSHNYVGGGNGIWNKLDNLYGLMTEYNYGSTESRRYKQKXGHDQIVRI 369
Db 1226 -----ETYEYNNKNDV-----TKQKDEGNVTIDVADGLDASETDQSGKSSAAV 1271
QY 370 ERTYNNVHLTSECKQCGQVQIQTETAYVAILGHNFDSQPSQFQLPKTKTETWRSADNSY 429
Db 1272 YDKYGNQIQSSKDLASNLK-----DGSPEAKSGWNLTASKD----- 1311
QY 430 RSEITETTFDESIGNPLTKVKKTKQKILSPS-----THWEYPPAGEVDNCPPEPYGFTFR 485
Db 1312 RRTIS-VIADKSG-----VLSGSALEVLQSSTAGTDHGYSSATQVELEPNTVTLTG 1365
QY 486 FVKKII---QTPYDSEFKDQPEKIQY---RYSLIGSQSHVTLKIBERHYSATQLLNS-- 537
Db 1366 KIKTDLAKSRAYFNIDLRDKDKRIQWIHNEYSALAGKNDW---KQKITFTPANAGK 1421
QY 538 ---TLFQVNTDKSELGL-LQTECTKGNGKTVSVVHKFTYTKQDDTLQOS-HSITHD 592
Db 1422 AVVMEVDHDKDKGKAWFDEQVLEKEGVSSSNYPVQNSFTSATENMNVNVCASVDSEE 1481
QY 593 NFTIHRQVRSRYTGRFLFSDTDTKDIIVTQMSYKLGRLLTET-----L 635
Db 1482 GFNDVSLKAARTSASQAGSVTKQTVVLGQANDKPYVLTILTGMSKASSVKFTDEKDYSL 1541
QY 636 NSGTPYANTLYDYEL---NNLQDNRPPFVI----- 664
Db 1542 QANVTYADSGTIGYNAKPPSGTQBNRAAVVVPKPKINKVDISILFKQSGATGTWPFDDI 1601
QY 665 -----TTTVDNGQLRNEFDGAGRHSVQCLKSDGCKEYTHITHTQYDQGRHHT 714
Db 1602 RLIEGSLTKSYDSNGYVTKEDELGYATS---TDYDETK---KTSETDAKGEKTT 1654
QY 715 STY---SDYLINGRQOTDPKVLHSMKSYDNWG-QIANT-----HWSYGVSEKI-- 760
Db 1655 YTYQADQDLTNWLSNGTSILH---SYDKEGNEVSKTIRAGADQYVKEFYDVMGLVK 1709
QY 761 TVDPIITLTKQLOSNMNVOT---GKEVTTYPSQOPIQITLDFEAGHLQSHCHITLRD 816
Db 1710 TTDPLGNVLASEYDANSNLTKTISPNGNEV-----SLSYD 1744
QY 817 QWDRVRKETDALGOCTIYQYDYNRNVQITLPLDGTIVNRKVAPFTDITLDIRVNGISL 876
Db 1745 GTRVRKSKSYNGTEKYIFTYDKNGN-----EYSVNKEQN-----TT 1781
QY 877 GQOTFDGLSRLTQSDQGRVWAYTYSAGNDQCPSTVITPDGQFIHYQVQPELDDAVLQVA 936
Db 1782 KKRTFDNKNRULTELDTRGSGQWTYPSDSKLKTP-----SWIH-----G 1821
QY 937 SNEITQOFSYNPVTCALLAKVAEGOSLPIIYPSGRKQWENIDMKVSYLWTLRGLENG 996
Db 1822 DQKGNQFTYN-----KLDQMIENKUSTSYSDYDEN- 1854
QY 997 YTDLTGCTIKISRDTHGRVTQIKDSSIKTILNYDNLNRHIGSOVTDLATGHMLTTTFEFD 1056
Db 1855 -----GNVQ-----TFITNGGGTFSYSDERNLVSLLHIGKXNGGDIETSEYV- 1898
QY 1057 GLNREIGRKLCDSSGHTLDIQOSWLKTOQLANRIVKLNGLVQRTQOYSYDSRBNLQVCK 1116
Db 1899 -----DANGNRRTINSS-----ASGRV-----QYBYGKLNQLVK--- 1927
QY 1117 DGAEPCTDKYGHISVTONFTYDIYGNITACHTTFADGTED--HAPFFKANPTDPCQLTEV 1174
Db 1928 -----ETHEDGTVI---EYTYDGFGRKTV-TTIKDGSSKTVNASFNIMN-----QLTKV 1973
QY 1175 H-----HHPD-DMPDNI-----RLKYDAGRVINLTDHNG 1203
Db 1974 NDESISYDKNGNRTSDGKFTYTWDAEDNLTAVTKKGDKPFATYKYDEKGNRIQKTVN-G 2032

QY 1204 NTEFTYDTLGLRQNGQSVGYDPLNRLVSKQTTDLCELY-YRETMVNEVRNGEMIR 1262
Db 2033 KVTNIFYDG-----DSLNVLYETDADNNVTKSYTYGD-----SGQLLS 2070
QY 1263 LIKRTGETTIIAQRASKVLLTGTDSQSVILTSQKONLSQAEYSAYGK-HKSTANDA---S 1318
Db 2071 YTEKCKYFYHYNAHGDIIAISDSGTQTV-----AKYQYDAWGNPTKTEASDEVKDN 2122
QY 1319 ILICYNGERADPVSGVTHLNGVRSYDPTLMSPHT--PDSLSPFGAGGINPYSYCLGDDPIN 1376
Db 2123 RYRYAGYQYDEBTGLYLLMRY--YEPNNGVFLSLDPPGSDGSDLDQNGYAYGNNNPVM 2180
QY 1377 RSDPSCGLISWQAWTIGMGIAGLLTIATGGMATAAAG-----GIAAAIASSTTTALAFCA 1432
Db 2181 NYDPPGH--W-VMLVNVNAGFA-----AVDGYKAYKSGKGMKGAWAAASNPFGPKIFKG 2231
QY 1433 LSVTSDIITSIVSGALEDASPKASSITLQVWSGM-----CAAGLAESAIGKGTKLATHLGAF 1488
Db 2232 ASRAYKFTK-----KAVKITGHTRHGLNOSIGRNG-----GRGVNLRKALNA- 2273
QY 1489 AEDGENALIKSTSESRIKWGVTRSLDRE--IVRNEEQVINKHSRGYTDNFMKGEOAI 1546
Db 2274 -----VRSPEKVIKQPNGATKYVKKATVVLNKGKVIYAG-----SS 2312
QY 1547 LVHGDQKGLYHTEGK 1563
Db 2313 RAKGSKHVEHTHGKNGK 2329

RESULT 6

ABU18641
ID ABU18641 standard; protein; 2234 AA.

XX AC ABU18641;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #4168.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Bacillus anthracis.

XX FN WO20027183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 08-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;

XX XX WPI; 2003-029926/02.

XX XX N-PSDB; ACA22511.

XX XX New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids, required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 46565; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are:

(1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 2234 AA;

Query Match
Best Local Similarity 4.3%; Score 386; DB 6; Length 2234;
Matches 339; Conservative 236; Mismatches 633; Indels 608; Gaps 76;

QY 21 FTQANNFTSAVSGGVDPDR-----TGLYNIQITLGHIVG-----NGNL-----57
DB 748 YSKVGHAESEISHEPVPFSDQSGFIGMFDYWSIPVLNGKVNATNGFMSEKIDITLSG 807
QY 58 -CPTLPPLTSLYSPNKTDIGFIGNFGLSV---YDRKNSLLSL-----TGEN- 102
DB 808 RPDVSVERTYNSQSKVGLFGTGWSSGLEERWADGNGNLLISTDGANITFRIGDNK 867
QY 103 -----YKVIETDKTKVLOOKLNLRFKDLKENCYRIIHKSGDIEVL 145
DB 868 YQAPGTIYLEIKQVSGGVEIKDKQTV-----TFYKSGDAQGR 905
QY 146 TGFNNNAEDLQVP-----KLLNPAGHAIYIDNFEATQPLNRIYDDLDGHDPL 196
DB 906 IEYTKDKYGNNTTVEYDGASLSKVKNASGKBLV---QYDGNKKAAARVI---GPDNKT 959
QY 197 LNLQYQGLIKTILTFPGQ--KEGYTELRLFLNQLNSIHNFSLGNENPLTWSFGY---250
DB 960 ITENYDGLLVSSITPEGVKXYGD-----NGVLTSLYDPQHTDAKPKTSYAVENDR 1013
QY 251 -----TPIGKNGIL-----GOWITSNAPGLKETVYNNQNGHHFPOSANLPLPVTL 301
DB 1014 LVKYTDLPLGRATLLAYNTGSKVELTTPKGRKTVTYND-----AGNFV-----1057
QY 302 MKQVPGAGQAPAIQAEYSVTSNHYVGG-----GSGIWNKKLDNLYGLMTEYN- 348
DB 1058 -KTVEDVGRNLNLTTSYEYNNANLVKTTPKNQETATYDNGNVTSDTDEMGTEKEFYNK 1116
QY 349 -YGTGERRYKDKB-----GHDQIVRIERTVNNVHLLTSECKQNGYIQTETAYYAI 400
DB 1117 DNGIIRKATDNEDRKTIVAYVGAITEVSQTDQGAN-----TSSVIHHDQYGNPIETSKELS 1171
QY 401 IGHNFDSQPSQFOLPKTKTETRSADNSYSRSEITETTFDESNGPLKVKDKTKIISP 460
DB 1172 AGNLIQNPS-FEM--NGTEKNVKVDYNNSGSIS-----KDATPAGGLGESSLKITTK 1223
QY 461 STH--WYYPYPAGVDNCPPEPVGFTFRVKII-----QTPYDSEFKDDP 503

DB 1224 ATNNDWGYIAAQEVTLPEPNTVTLTSGWVKTLVNGAFAFNVSQSLNENGAGIDGGWHDTR 1283
QY 504 EKFIQVRYSLIGSOSHVTLKIEERHYSATQLLNSTLFQYNTDKSELGRU-LKQTECTKE 562
DB 1284 HNKVOGTSDWVNRQ---VTRKTEQ-----TRKVIYLOVENGSATSGSAWFKIOLKEKE 1337
QY 563 NGKTSYVHKFTYTKQ--DDTLQQ-SHSITTH-----DNFTIHRSVORRYTGRLEFS 611
DB 1338 VSSFPVLNSPEENPDGFPQWVRSCQCHERNVDSDSFTGHSSIVMER-----S 1391
QY 612 DDTKDKIVTQMSYDKLGRLLTRILNSGTYPANTLYDYELNLODDNRPPFVITTDVNG 671
DB 1392 EYGPNDI-----GYRNRVILNQKAEVTLTAMSKSENVND-----1428
QY 672 NOLNREFDAGRHVSQCLSDGDFYTHIQDYDEQCRHHTSTVSDYLTNGRQOTDP- 730
DB 1429 -----APDKLSKDYAVLAETYYQD-----TVVNYTSFPGSTNDW 1464
QY 731 -----DKVHLSMSKSYONWGOIANTHWSYGVSEKIVDPITLTATKOLQNS 777
DB 1465 NRSAAVIPAKKPIQKIEIFLFRKNKG-----KVPFDDIRLLEGNALIKNE 1511
QY 778 NNVTGKEVTTTPSQOPIQITLDFEAGHLOSCHTLTRGMDVRKETAIGQCCTIYQD 837
DB 1512 YD-NDGNVATY-----DEBQK---NTFTVDASGNKKSETDEKGNTKLYDYN 1555
QY 838 NNNRVITLPGTIVNRKIAPFSTDTLITIRVNGISLGQOTPDGLSRLTQSDGGRW 897
DB 1556 KDNLLTKVTLKNGTSVNVRYD-----1576
QY 898 AVTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLASNEITQOFVSNPVTGALLKAV 957
DB 1577 -----HNGNTEKSVWFGKTKTHKEYVDVNDKNTVYIDALARRIEN-TYDENANKIKTYM 1631
QY 958 AEGQSILPIYPSGLKMKMENINDMKNGSYLWLRGLENGYTDLTGTQIKISRDTHGRVTQ 1017
DB 1632 PNGSILESVDYDRAVVGEEK-----RNGKDSFT-----FERDQNGQVTK 1670
QY 1018 IKD--SSIKTILNYDDLNHRHGSQVTDLATGHMLTTTVEFDGLNREIGRKLCDSSGHTLD 1075
DB 1671 VKDLVNGVERTKTYDKADR-----VTS--AT-----DSRGGKID 1702
QY 1076 IOQSWL-----KTQQLANRIVKLNGVLQRTQEQSYSDSRNLNQ--YKCDGAECPDXY 1126
DB 1703 -----NAYHDKANSKTEKLEKQIVTQGY--TNKVSYD-YNTLDQNIIVTDSQ-- 1748
QY 1127 GHSIVQNTFTYDIYGNITACHTTFADGHEDHATKFF--ANP-----TDFCQLTEVHHT 1177
DB 1749 -----TYRFDYDQGNV-----RTYTAGNGSGSTFNYDQANKIKOLVWGTNSILLSEYE 1799
QY 1178 HPDMPDNIRLKVDKAG-----RVINITDNHGNENTFTYDTLGR-----LQ 1217
DB 1800 YDQSNRRTKIKHEGAGGKVETNFVYDPIQLLNEVLNPGTTSYTYDGNRTSVKVIIE 1859
QY 1218 NQGS-----VYGVDPNRLVSKQTDLDCEL-----1244
DB 1860 NGKETKSAATFNEGNQJLVKFGNESLTYDVANGNRTSDGKYKYTWNEDDQIVAITKQGENN 1919
QY 1245 -----YVRETMVNVNVRNGEMIRLLRTGETI-----IAQ 1274
DB 1920 AFATKYDEDNRREKQNVNGQVTRYFYDQDSINPLIYETDNGTGVLRQVYVSADARLAMK 1979
QY 1275 RASKVLLTGTSSQOSVI--LTSQKQNLQSOEAYSAGK-----HKSTANDASILYNGERA 1327
DB 1980 AQGQTLYYHNPGRGDVWAMTNDQKEVATYEDAWGNVLTSDTKGIAAD-NPFGYAGMY 2038
QY 1328 DVSQVTHLNGYRSYDPTLMREHTPD-----SLSFPGAG-----GNPYXSYCLGDPINR 1377
DB 2039 DKEIGMYL-----IARYNPEHGVFLSDPDPGDEDDPVTNGTYADNPNVMM 2088
QY 1378 SDPSGHLISQAWTGIQMGIAGLLTIATGGNAIAAGGIAAAIASTSTTALAFQALSVTS 1437
DB 2089 TDPGKMAW-----LVPVVVIAGMVAARFCAKYAI-----RYGAKYKGVKAVKS 2131

QY 1438 DITSIVSGALEDASPKASSILGWYSGMGAAGLAESAIAKGGTKLATHLGAFAEDGENALL 1497
 Db 2132 -----GWY-----GKVAKSGWNKGSIAQKIPRIHKVGR---I 2163
 QY 1498 KSTSESRIRKMGVTRSLDREIVRNEEGQVIKDHSGRYTDNFMGRGEOAILVHGDGKDGFLY 1557
 Db 2164 KGDNDKGYWGVYISTK-----KTGRKTY-----SFEFHPHNGHY 2203
 QY 1558 HTEGNKNG-KGPYTR 1572
 Db 2204 HLQNKYSKYQGKWR 2219

RESULT 7
 ADC01365
 ID ADC01365 standard; protein; 1400 AA.
 XX
 AC
 XX
 AC
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE Enterohaemorrhagic E. coli O157:H7-specific protein SEQ ID NO: 1410.
 XX
 KW enterohaemorrhagic; anti-bacterial.
 XX
 OS Escherichia coli; O157:H7.
 XX
 PN JP2002355074-A.
 XX
 PD 10-DEC-2002.
 XX
 PF 24-JAN-2002; 2002JP-00015959.
 XX
 PR 24-JAN-2001; 2001JP-00112010.
 XX
 PA (UUTS-) UNIV TSUKUBA.
 XX
 DR WPI; 2003-451640/43.
 XX
 PT Enterohaemorrhagic Escherichia coli O157:H7-specific nucleic acid molecule
 PT and a polypeptide and its use, a polypeptide, a vector and a host cell.
 XX
 PS Claim 3; SEQ ID NO 1410; 2067pp; Japanese.
 CC
 CC The invention relates to a novel enterohaemorrhagic Escherichia coli
 CC O157:H7-specific nucleic acid molecule. A polynucleotide of the invention
 CC has anti-bacterial activity. The polypeptide can be used in detection
 CC and/or treatment of O157:H7 infection. The nucleotide sequence of the
 CC genome of Enterohaemorrhagic E. coli O157:H7 was determined. The present
 CC sequence represents an E. coli O157:H7-specific polypeptide of the
 CC invention.
 CC
 SQ Sequence 1400 AA;

Query Match
 Best Local Similarity 4.1%; Score 365; DB 7; Length 1400;
 Matches 284; Conservativity 20.9%; Pred. No. 7.5e-16; Mismatches 467; Indels 456; Gaps 65;

QY 418 KTEW-READNSYRSEITETTFDSGNPLTKVI-----XDKTKIISPSHWEY 466
 Db 218 RLTIVRREAAGDLAGEITGVT-DGAGREFRLVLTQAQRAEARKQHTASLSSPDT--- 272
 QY 467 YPAGEVNCPEPYGFTFRVKIIQTPYDSEFDKDEKFTQYRYSLIGSOSHVTLKIEE 526
 Db 273 -----PRPLSDSAFPDTPUGTEY-----GPDGRGR----- 297
 QY 527 RHVSATQLNLTFOYNTDKSELGRLLQTECTKGNGKTYSV-----VHKFTYTKQD 579
 Db 298 --LSAVNLTHPAPEPESLPGLARY-----TYTEAGELLAVYDRSNTQVRAFTYDAQH 349
 QY 580 DTLQOSHITTHDNFTIHRSQVRSRYTGRILFSDTDKDIQTOMSYDKLGRLLTILNSGT 639

Db 350 PGRWVAH-----RYAGR-----PEMYRYDDTGRVVVQLNPAGL 383
 QY 640 PYANILTYDYELNLIQDNRPPFVITTTD-VNGNQLRNEFDGAGRHSQCLSDSGDGKF 698
 Db 384 SY-----RYQYE-----ITVTDLSLRREVLHTEGGAG-LKRVRKKEADG-- 425
 QY 699 YTIHTQYDEQGRHHTSTYSYDYLITNGR--QOTDDPKRVHLSMSKSYDNMGQIANTHWSYGV 756
 Db 426 -----SVTHSGYDAAGRLTAQTD-----AAGRTEYGL 453
 QY 757 SEKITVDPIITLTKQLQNSNNVQTGEVITYTTSQQPIQITLDFEACHLOS-----CHT 812
 Db 454 N--VWSGDITDITPDGRETKFYNDGNQLTAVSPDGLSESRAYDEPGRVSETSRQGD 511
 QY 813 LTRDGWDRVRKE-----TDAIGOCITIOYDNNRVNIQITLPDGTIVNRKYAPFSDTLIT 867
 Db 512 VIRAYDNPHSELPAITTTDATTSTQMTWSRGQLLAFTDCSGYQTRVEYDFRGQMTAVH 571
 QY 868 DIRVNGISLGQOTFDGLSRLTQSDQG-GRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQP 926
 Db 572 --REGIS-RYRRYDNRGRLTSVKDAQGHETREYENAAAGDL-TAVITPDGNSRSETQY-- 624
 QY 927 ELDDA---VLQVASNEITCOQPSYNPVTGALLKAVAGQSGLTPIYYPSGRKLKMINDMKK 983
 Db 625 ---DAWGKAVSTTQGGLTRSMEYD-LAGRITTLTNGSRSEPTYDA----- 667
 QY 984 MSYLWTLRGLNGYTDLTCTIQKISRDTGRVTOIKDSSIKTLLNYDILNRHIG----- 1037
 Db 668 LDRLVQQRGFD-----GRTQRYHYDLTKLTQSEDEGLVTLWHYDESDRLTHRTVNGE 720
 QY 1038 -----SQVTDLATGHMLTTTVEFPGLNREIGRKLCDSSGHTLDI-----QOS 1079
 Db 721 PABQWOYDEHGLWTEISHLSEGHQVAVHYGYDKGRLAGERTVHNPFETGELLMOHETSH 780
 QY 1080 WLKTOQLANRIVKLVGLQRTQYSYDS-----RNRL-----N 1112
 Db 781 AYNEQGLANRVTP--DSLPRVWLTGYSGYLAGMKGGTLPVFEFTDRHLREIVRSFGNN 838
 QY 1113 QYKCDGAECPDKYGH-----SIV-----TONFYDIYGNIT 1144
 Db 839 AVELTSTYTPA---GHLQSORLNSQVYDRDYDNNDGDLVRISGRQTWEYGYSATGRLE 895
 QY 1145 ACHTTADGTEDHATFKFANPTDPC--OLTEVHHTHPD-----MPDNRL-----KY 1189
 Db 896 SVKTLASD-----LDIRIPATDPAGNRLPD-PELHPDSTLTAWPDN-RIADAHVYVYH 948
 QY 1190 DRAGRVINITD-----NHGNTENFTYDTLGR-----ONGGSV---YGYDPL-- 1229
 Db 949 DEYGRUTEKTDRTIPAGVIRTDDBERTHHYDSSHRLVYFTRIQHGEPLVESRYLYDPLGR 1008
 QY 1230 -----NRLVSQKTDITLDCELYR----- 1247
 Db 1009 RMAKRVRRERDLTGWMSLSRKEPVWYGDGDRUTTVQTDTRITQTVVPEGSFTPLIRV 1068
 QY 1248 -----ETMLVNEVRNG-----EMIRLRTGETIIAQRAS----- 1277
 Db 1069 ETENGEBREKAQRSLEAETLQQEGSENGHGVVFPAPAEIVRLDLREBEIRADRVSSESRAWL 1128
 QY 1278 -----KVLVTGTDQO--QSVILTSQKMLSQEAYSAYKHKST 1313
 Db 1129 AQCGLTVEQLARQVEPEYTPARKVHFYHCDHREGLPLALISEDGNTAWRGEYDEWGNQLNE 1188
 QY 1314 ANDASI---LGVNGERADPVSGVTHLNGYRSGYDPTLMRFHTPDSISLSPFG-AGGINPYSY 1369
 Db 1189 ENPYLHQYRPLPGQHQDEESGLYNNRY--YDPLQGEYITQD---PIGLAGWNLYNY 1243
 QY 1370 CLGDPINRSDPSG-----HLSQWNTGICNGIAGLLLT-----IATG---GMAIAAGGI 1416
 Db 1244 PL-NPIIRMDPLGLNLYQLLYDVVHDDSYGTSIDITGSGDLISLGGHAGLGAFAPKK 1302
 QY 1417 ARAIASTITALAFGALSVTSDITSVSGALEDASPKA--SSILGWVSMGMGAAGLAESA 1474
 Db 1303 GEMLSIDICIYATACGAGIGGGINAAITYSKSLPTSGVNSVG-VTVGGGVGG----- 1356

749 GNLTHEKTRPVQQLRFGQVLDRETGLHY-
NLYRFYDDPDGKFKISGD-
---PIGLAGG 803

1364 INFYSYCLGDPINRSDPSG 1382
||| | : : :
804 INLYQYA-ENPLSYIDPLG 821

RESULT 9
ABU19676
ID ABU19676 standard; protein; 1515 AA.
XX AC AC
XX ABU19676;
XX DT
XX 19-JUN-2003 (first entry)
XX DE
XX Protein encoded by Prokaryotic essential gene #5203.
XX DE
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX KW
XX Borrelia cepacia.
XX OS
XX PN WO200277183-A2.
XX PD
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX PR
XX 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA
XX (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX DR N-PSDB; ACA23546.
XX PT
XX New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PS
XX Claim 25; SEQ ID NO 47600; 1766pp; English.
XX CC
XX The invention relates to an isolated nucleic acid comprising any one of
XX CC the 6213 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX CC polypeptide or its fragment whose expression is inhibited by the
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX CC proliferation or the activity of a gene in an operon required for
XX CC proliferation; (7) identifying a compound that influences the activity of
XX CC the gene product or that has an activity against a biological pathway
XX CC required for proliferation, or that inhibits cellular proliferation; (8)
XX CC identifying a gene required for cellular proliferation or the biological
XX CC pathway in which a proliferation-required gene or its gene product lies
XX CC or a gene on which the test compound that inhibits proliferation of an
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX CC compound's activity; (11) a culture comprising strains in which the gene
XX CC product is overexpressed or underexpressed; (12) determining the extent
XX CC to which each of the strains is present in a culture or collection of
XX CC strains; or (13) identifying the target of a compound that inhibits the
XX CC proliferation of an organism. The antisense nucleic acids are useful for
XX CC identifying proteins or screening for homologous nucleic acids required
XX CC for cellular proliferation to isolate candidate molecules for rational
XX CC drug discovery programs, or for screening homologous nucleic acids
XX CC required for proliferation in cells other than *S. aureus* *S. typhimurium*,
XX CC

Db 1225 RDG-----TTGQVVTWHIEPGSFLPLAQETDDGLPILTDQIGRPKTVDEQGRPVWKA 1278
 QY 1303 AYSAYGK---HKSTANDA-----SILYNGERADPVSGVTHLNGYRSYDPTLMRPH 1351
 Db 1279 AYLWGLKLLPKVRPANDACGATSIDTILRSGGWADDETGLNYLNRY--YFDSQYL 1336
 QY 1352 TPDLSLSPRG-AGGINPVSYCUGDPINRSDPG 1382
 Db 1337 SAD---PIGLGGGARTOAY-VHDPESQWIDPLG 1364

RESULT 10

ABU15135
 ID ABU15135 standard; protein; 1397 AA.
 XX
 AC ABU15135;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #662.
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS Escherichia coli.
 XX
 PN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlson KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamanoto R, Forsyth RA, Xu HH;
 XX
 DR WPI; 2003-029926/02.
 DR N-PSDB; ACA19005.

New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

Claim 25; SEQ ID NO 43059; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1397 AA;

Query Match 3.9%; Score 342; DB 6; Length 1397;
 Best Local Similarity 21.4%; Pred. No. 3.1e-14;
 Matches 256; Conservative 157; Mismatches 382; Indels 404; Gaps 63;
 QY 418 KTEW-RGADKSYRSEITETTFDESGNPLTKVI-----KDKTKKIIS-----PS----- 461
 Db 218 RTQTHREAGFEISGITGVT-DGAGRHERLVLTQAQAEAEARQAISGCTESAPDPT 276
 QY 462 -----THWEYPPAGEVDCNCPPEP---YGFT-----RPFV 488
 Db 277 LFQYTEYGRDNGIRLSAVMLTHDPEYP-----ENLPAAPLVRYGWTPRGEAAVYDRSNT 331
 QY 489 KLIQTPYDSEPK-----DDPEKFTQYRSLIGSOSHVTLEERHYSATQLLNSTL 539
 Db 332 QVRSFYDDKYRGVMVAHRHTGRPE--ICRYD--SDGRVTEQLNPAGLSVT----- 379
 QY 540 FOYNTKSELGRLLKQTEC--TKGNGKTYSVVHKFTYTKODDTLQOSHSTTHDNFTIH 597
 Db 380 YQYKDRITITDSLNRREVLTQEGG-LKRVRK-----EHADGSS 421
 QY 598 RSQVESRYTGRLFSDTDKDIVTQMSYDKLGRLLTRTLNSGTPYANTLYDYELNNLQDD 657
 Db 422 QSQFDA--VGLRAQTDAGRTTEYSPVVTGLTRIT---TPGRASAFY----- 468
 QY 658 NRPPFVITTDVNGNQLRNEFDGAGRHYVSCQLKSD-----GDGKFYTIHTQQYDEQGRH 712
 Db 469 NHHSQTSATGPDGLEIRREYDEWGRLIQETAPDGDITRYRDNPHSDLPATEDATGSR 528
 QY 713 HFTSYSDVLTNGRQQTDPKVLHLSKSYDNWQLANTHMSVGYSEKITVDPI-TLTATK 771
 Db 529 KTMWSRY---GQLLSFTDCSGYVTRYDHRFGQVAVHREELGSLQYRAYDSRQLIAVK 585
 QY 772 QIQSNNNVQTGKVTYTPSQQPIQITLFDAGHLQSCHTLTRDGMWRVAKETDAIGQ- 830
 Db 586 DTQGHETRYE-----YNAAGLTT--VIAPDG-SRNGTQYDAMGKA 623
 QY 831 CTI-----YQYDNNYRVIQITLPDGTIVNRKYAPESTDTLITDIRVNGISLQQTFF- 881
 Db 624 ICTTGGLTRSMEXDAAGRVIRLTSENGS-----HTTFRYDVLRLIQTGDFDQRTQYH 678
 QY 882 -DGLSRLTQSDQGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVANSNEI 940
 Db 679 HDLTCKLIRSEGLV-----THWYD-EADRLTHRTVNGET 714
 QY 941 TQOFSPYNPTGAL--LKAVAEGSLTPYI-YPS-GRLLXENI-----NDM---KMSY 986
 Db 715 AERWQYDS-RGWLTDISHISEGHRVTVHYGYDGSGRASEHLTVHPQTNELLMQCHETRH 773
 QY 987 LWTLRGLNGVYDLTGTTQKTSRDTGHRVQTQKSSIKITLNYDDLNRHIGSQVTDLATG 1046
 Db 774 AYNAGLAN--RCIPDSLPVAVENLTYG-----SGWLSGMKLG 808
 QY 1047 HMLTTVTTF--DGLNRETRGRKLCDSSGHTLIDIQOSWLKTOQLANRIVKLVGLORTQYS 1104
 Db 809 D--TFLVSYTRDLRHRETLR-----SFGVELTAYTPAGQLQSQ--HLNLSLS-DRDYT 858
 QY 1105 YDSRNLRLQYKCDGAECPTDKVGHISIVTONFYDIYGNITACHTTTFAD-----GTEDHA 1158
 Db 859 WNDNGELIR-----ISSPRQ-----TRSYSYSTTGELTGVHTTAANLDIRIPYTTDPA 906

QY 1159 TFKFANPTDPCQLTEVHHTHPD-----MPDN-----IRLKYDKAGRVINITD----- 1200
Db 907 GNRLPDP-----ELHPSALSMMWPDNRRIADAHYLYRYDRHGRLEKIDLLIEGV 956
QY 1201 ---NHGNTENTYDTLGRLOQSGSVG-----YDPLNRVLSQKTDLDEL----- 1244
Db 957 IRTDDERTHRYHDSQHLVHTVTOYAEPLVESRYLYDPLGRVRAKRVRRERDLTGM 1016
QY 1245 ---YV-----RETLVNE-----VRNGEMIRLLR-----TGETIIAQORA--- 1276
Db 1017 SLRKPQVWYWGWDGDRUTTIIONDKTRIQTITYQPSFTPLIRVETATGELAKTORRSLAD 1076
QY 1277 ---SKVLLTGTDSQQSVILTSKQMLSOEA----- 1303
Db 1077 TLQSGGSDGGSVVFPVVLQMLDRLESEILA---DRVSEESRWLASCGITVAQMOSQM 1133
QY 1304 ---YSAVKKHKSSTANDAS-----ILGYNG 1324
Db 1134 DPVTPARKIHLXCHDRHGLPLALISTEGTTAWYAEYDEWGNLLNEENPHLOQLIRLPG 1193
QY 1325 BRADPVSGVTHLGNRGYSYDPTLMRPHFPDLSLSPFG-AGGINPYSYCLGDPINRSDPSG 1382
Db 1194 QQYDEESGLYY-NHRHYVDPLQGRYITOD---PIGLKGGWVFOYPL-NPISNIDPLG 1246

RESULT 11
ID ADC00960 standard; protein; 1404 AA.
AC ADC00960;
XX
DT 04-DEC-2003 (first entry)
DE Enterohaemorrhagic E. coli O157:H7-specific protein SEQ ID NO: 1004.
KW enterohaemorrhagic; anti-bacterial.
XX
OS Escherichia coli; O157:H7.
XX
FN JP2002355074-A.
XX
PD 10-DEC-2002.
XX
PF 24-JAN-2002; 2002JP-00015959.
XX
PR 24-JAN-2001; 2001JP-00112010.
XX
PA (UYTS-) UNIV TSUKUBA.
XX
XX WPI; 2003-451640/43.
XX
PT Enterohaemorrhagic Escherichia coli O157:H7-specific nucleic acid molecule
PT and a polypeptide and its use, a polypeptide, a vector and a host cell.
XX
PS Claim 3; SEQ ID NO 1004; 2067pp; Japanese.
XX
CC The invention relates to a novel enterohaemorrhagic Escherichia coli
CC O157:H7-specific nucleic acid molecule. A polynucleotide of the invention
CC has anti-bacterial activity. The polypeptide can be used in detection
CC and/or treatment of O157:H7 infection. The nucleotide sequence of the
CC genome of Enterohaemorrhagic E. coli O157:H7 was determined. The present
CC sequence represents an E. coli O157:H7-specific polypeptide of the
CC invention.
XX
SQ Sequence 1404 AA;

Query Match 3.8%; Score 341.5; DB 7; Length 1404;
Best Local Similarity 20.2%; Pred. No. 3.4e-14;
Matches 255; Conservative 130; Mismatches 370; Indels 505; Gaps 59;
QY 376 YHL-LTSECKQONGYIQTETAYVAILIHNFDSPSQFQPKT-KTETWRSADNSYRSEI 433
Db 245 FHLVLTQAKRAEVQRQRATSLSSPAGPR--SASSSLVFPDITLPAGTEYGADNGIRLEA 302

QY 434 TETTFDES-----GNPLTKVIKDKTKQKIISPSTHWEYYPAGEV-----DNCPPPEYCF 483
Db 303 VMLTHDPAYDELPAAPLARYT-----YTASGELRAYVDSRGTVQVRGF 345
QY 484 TRFVKKIIQTPYDSE-----FKDDPEKPIQYRSLIGSQSHVTLKIEERHYSATQL 534
Db 346 A-----YDAEHAGRWAAHVHAGRPES--RYRYDDTG-----RVTEL 379
QY 535 LN-----STLPOYNTDKSELGRLLKQTEC--TYGKNGKTVSVVHKTYTKQDDTLQOQSHSI 588
Db 380 VNPEGDLYRPEYQGRVITITDSLNRRREVLYTEGEG-LKRVVVK----- 422
QY 589 TTHDNFTIHRQSVRSRYTCGRLEFSDTDTKDIVTQMSYDKLGRLLTTLNSGTFEVANTLYD 648
Db 423 -EHADQSITRSE-----YDEAGRLKACT----- 444
QY 649 YELNNLQDNRPFVITTTDVNGNLNRFDFGAGRHSOCLKDSG-----DGKFTY 700
Db 445 -----DAAGRTTEYSLHMASGAVTAVTGPDR--T 472
QY 701 IHTQQYDEQGRHHTSTYSYDLTNGRQQTDPDKVHLMSKSYDNWGQIANTHWSYGVSEKI 760
Db 473 VR-YGNSORQVTSVTPDGLRSSREYDEKGLAAETSRSGE-----TTRYSD----- 520
QY 761 TVDPITLTATKQLQSNNNVQTKVITYTPSQPIQITLFDGAGHLQSCHTLTRDGWDR 820
Db 521 --DP-----ASELPTGIQDATGSTRQM-----AWSR 544
QY 821 ---VRKETDAIGCTIYQYDNNRVNVIQITLPGTIVNRKYAPFSTDTLITDIRVNGISLG 877
Db 545 YGQLLTFDTCGTYTRYEDRYGQGIHVHREGLISTYSSYN-----RG 588
QY 878 QQTFDGLSLRSLTQSDQGRVWAYTYSAGNDQCSTVITPDGQIHYQYQPELDDA---VLIQ 934
Db 589 Q-----LVSKQDAQGRETRYEYSAAGDL--TAIVAPDGSSEIQY-----DANGKAVS 634
QY 935 VASNEITQOFSYNPVTGALLKAVAGOSLTPYYPGSRGLKMNINDMKKVSLWTLRGLE 994
Db 635 TTQGGLTRSMGYDAAGRITVLTNENGSSQSTFRYDPRDLTEQGRFDGRTQRY----- 686
QY 995 NGYDLTGTI-----OKISRDTHGRVTOIKDSS-- 1022
Db 687 ---HYDLTGKLTQSEDEGLVTLWHYDASDRITHRTVNGDPAEQWQYDEHGWLTLSHTSEG 744
QY 1023 ---IKTTLNVDLNRHIGSQ--VTDLATGHML----- 1049
Db 745 HRVSVHYGYDDKGRLLTGERQTVENPETGEMLWEHETGHAYSEQGLATROEPDGLPVEWL 804
QY 1050 -----TTTVEF--DGLNREIGRKLCDSSGHTLDIQOSWLKTCQLANRIVKL 1093
Db 805 TYGSGVLAKMGLGGTLPVEMDRDLHRETPARF--GGEAYELATAWNTSQGLRSRHLML 861
QY 1094 NGVLRQTEQYSDSNRNLNOYKCDGAECPDKYGHISVTQNFTYDIYGNITACHTTFADG 1153
Db 862 P---QLDRDYDWDNGQL--IRISGPQ-----ESREYRSDTGLTGVHTAANL 906
QY 1154 TED--HATFKFANP-TDPCQLTEVHHTHPD-----MPDNRL-----KYDKAGRVIN 1197
Db 907 DIDIPYATDPAGNRLPDP-----ELHPDSTLTAWPON-RIAEADAHVYRYRYDEYGRLEA 958
QY 1198 ITD-----NHGNTENTYDTLGR-----QNGQGSV--YGYDPL----- 1229
Db 959 KTDRIPGEGVIRMHEDERTHHYVDSQHLRVFHTRIQHGEPOVESRYLYDPLGRRTGKRVWR 1018
QY 1230 -----NRLVSQKTDITDCELYR-----ETMLVNEVNGEMI 1261
Db 1019 RERDLTGWMSLSRKPEETWYWGWDGRLTIVTQQTTRIQTIVYQPSFTPLRIETENGQA 1078
QY 1262 RLLR-----TGETIIAQ-----ORASVLLTGTDSQOS-----VILTSPKQNL 1300
Db 1079 KAHRSRLAEVLQEDTGVTLPAELAVMLGRLERLRQGSVSESQWLACGLTAECMQAQ 1138

QY 1330 VSGVTHLNGVRSYDPTLMRFHTPDSLPFG-AGGINPYSYCLGDPINRSDPSGH-SWOA 1388
 Db 1199 ESGLYY--NRHRYDPLQGRVITQD---FIGLKGWNLGYQL-NPISDIDPLGLSNWED 1252
 QY 1389 WTGIGMGIAGLLLTATGGMALIAAAGGIAAAIASTSTTALAFGALSVTSDITISVGALE 1448
 Db 1353 -----AKSG---ACTNGLCGTLSA-----MIGPKFDSIDSTAY 1283
 QY 1449 DASPKASSILGWSMGMAAGLAESAIAKGGTKLATHLGAFAEDGENALLKSTSESSRIKW 1508
 Db 1284 DALNKINS-----QSICEDKEFA-----1301
 QY 1509 GVTRSLDREIVRNEGQVIKHSRGYTDNFMWGEO-----AILVHGDKDGLFVHTEG 1561
 Db 1302 -----GLICKDNGRYFSTAPNKGKSGSYFPNPGNGTEKVSAYHTHG 1346
 QY 1562 NKHNGK--GPYTRHTPEQLVDYLKDNKI 1587
 Db 1347 ADHGEYWDIEFSGKDEKIVK-SKDNKI 1373

RESULT 14

AA15983
 ID AA15983 standard; protein; 1426 AA.

AC AA15983;

DT 05-OCT-2000 (first entry)

DE E. coli proliferation associated protein sequence SEQ ID NO:340.

ES Escherichia coli; E. coli; proliferation; inhibition; screening;
 KW antimicrobial; bacterial growth; antisense therapy; antibacterial.

OS Escherichia coli.

FN WO200044906-A2.

PD 03-AUG-2000.

PF 27-JAN-2000; 2000WO-US002200.

PR 27-JAN-1999; 99US-0117405P.

PA (ELIT-) ELITRA PHARM INC.

PI Zyskind J, Ohlsen KI, Trawick J, Forsyth RA, Froelich JM;
 PI Carr GJ, Yamamoto RT, Xu HH;

DR WPI; 2000-514822/46.

DR N-PSDB; AAA65988.

PT Novel polynucleotides and polypeptides associated with microorganism
 PT proliferation, used to identify inhibitors of bacterial growth and
 PT proliferation, for use in antisense therapy.

PS Claim 11; Page 253-256; 316pp; English.

CC AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide
 CC sequences derived from *Escherichia coli* which inhibit *E. coli*
 CC proliferation. AAA65890 to AAA66055 and AAA65886 to AAA66040 represent
 CC nucleotide and protein sequences associated with *E. coli* proliferation.
 CC AAA66056 and AAA66057 represent primers used for sequencing *E. coli*
 CC proliferation inhibiting nucleotide inserts in an example from the
 CC present invention. Methods from the present invention can be used to
 CC identify a proliferation- required gene in a microorganism, by contacting
 CC a microorganism with a proliferation- required gene activity inhibitor
 CC nucleic acid identified in another organism, and determining if
 CC inhibition occurs in the second microorganism. The nucleic acid sequences
 CC identified as being required for bacterial growth and proliferation, can
 CC be used for antisense therapy for killing bacteria

SQ Sequence 1426 AA;

Query Match 3.8%; Score 339; DB 3; Length 1426;

Best Local Similarity 20.2%; Pred. No. 5.2e-14;
 Matches 257; Conservative 150; Mismatches 409; Indels 458; Gaps 60;

QY 572 KFTYTKQDDTLQOQSHSIITHDNFTIHRSQVRSRYTGRSLFSDTDTKDIVTQMSYDKIGRL 631
 Db 317 RYTYTEAGELL-----AVYDSNTQVRAFTYDAQHPGRVAHRYAGRPMPRYRYDDTGRVV 372
 QY 632 TETLNSGTFYANTLTIDYELANLQDNRPPFVITTTD----- 668
 Db 373 EQLNPAG-----LSRY-----IIVTDSLNRREVLHTEGAGLKVVKKE 418
 QY 669 -VNGNOLNEFDGAGRHSVQCLKSDGDGKRYTHITQOYD-----EQGRHHSTYSDYL 721
 Db 419 LADGSVTRSGYDAAGRLTAQ--TDAAGRRTEYGLNVVSGDITDITTPDGRHETKRYND-- 474
 QY 722 TNGROQT---DPDKVHLSMSKSYDNWGQIANTHWSYGVSEKITVDITLTATKQLQNSN 778
 Db 475 --GNQLTAVVSPD--GLSRREYDEPGLV-----SE 502
 QY 779 NVQTGKEVT--TYTPSQOPIQITLFEAGHLOSCHTLTRDGDWRVRKETDAIGQCTIYYQD 837
 Db 503 TSSRGETVRYRYDDAHSELPAITTTDATG---STRQMTWSRYGQLLAFTDCSGYQTRYEYD 559
 QY 838 NVNRVIQITLPGDTIVNRKYAPFSTDTLITDIRVNGISLQOQTFDGLSRLTQSODG-GRV 896
 Db 560 RFGM-----TAVIRE-----EGISL-YRRYDNRGLTISVKDAQGRE 595
 QY 897 WAYTYSAGNDQCPSTVITPDGQFIHYQV----- 925
 Db 596 TRYENAAAGDL--TAVITPDGNRSETQYDAMGKAVSTTQGLTRSMEDYDAAGRVISLTNE 653
 QY 926 -----PELDDAVLQVANSNEITQOFSYNPVTGALLKAVAEQSLTPYIPSGRLKME 976
 Db 654 NGSHSVFSDALDLRVQGGGFGRTQRYHYD-LTGKLTQSEDEGLVILWYDESDRITHR 712
 QY 977 NINDMKMSYLWTLRGLNGEYTDLTGTIQKISRDRTHGRVTQIKDSS-----IKTTLNYDDL 1032
 Db 713 TVNGEPAPQWQY-----DGHGWLTDISHLSEGHRAVAVHYGDDK 751
 QY 1033 NRHTG--SQVTDLATGHL----- 1049
 Db 752 GRLTGECQTVENPETGELLWOHETHYANEOGLANRVTPDPSLPVEMVTGSGYLAGMKL 811
 QY 1050 -TTTVEF--DGLNREIGRKLCDSSGH--TLIDIOQSWLKTQQLANR----- 1089
 Db 812 GGTPLVEYTRDLRHRETIVRSFGSMAGSNAAYELTSTVTPAGLQSQHLNSLVYDRDYGS 871
 QY 1090 ----IVKLVGLQRTQYSYDSRNLNQYKCDGAE-----PTDKYGHST----- 1130
 Db 872 DNGDLVRISGRQ--TREYGSATGRLESVRTLAPDLDIRIPYATDPAGNRLPDPPELHPS 930
 QY 1131 -----VTQN-----FTYDIYGNITACHTTFADG--TEDHATFKFANPTDPCQLTEV 1174
 Db 931 TLTWVPDNRIAEDAHYVYRHDEYGRLETKTDRIAGVIRTDDET----- 975
 QY 1175 HHTHPDMPDNIRLKYDKAGRVINITD-NHGN---TENFTYDTLGR----- 1215
 Db 976 HHYH-----YDSQHLRVFYTRIQHGEFELVESRYLYDPLGRMAKRVWRERDLTG 1025
 QY 1216 ---LQNGQSGVGYDPLNRLVSKOTDITLDCELYR----- 1247
 Db 1026 WMSLSRKEFVWTYGDG-DRLTVTQTDTTRIQTVYEGSFTPLIRVETENGEREKAQRS 1084
 QY 1248 --ETMLVNEVANG-----EMIRLL-RTGETIIA-----QQRASKV 1279
 Db 1085 LAETLQEGSENGHGVFPFAPBLVRLDLRLEBEIRADRVSSERAWLAQCGLTVEQLARQV 1144
 QY 1280 LLTGTDSCQS-----VILTSKQNLQSOEA--YSAYGKHKSTANDASILCYN--- 1323
 Db 1145 EPEYTPARKAHLVHCDHRGLPLALISEDGNTAWSAEYDEWGNQNLNEENPHHV--YQPYRL 1202

Qy	1324	-GERADPVGWTHLNGVRSYDPTLMRFHTPDSLSPTG-AGGINPYSYCLGDPINRSDPS	1381
Db	1203	PGQQDEESGLYY--NHRHYVDPLQGRIYTOD---PMGLKGMWNLYQVPL-NPLQIQIDPM	1256
Qy	1382	GHLSSQAWTIGIMGIAGLLLTATGGMAIAAAGGIAAAIATASTTTALAFGLSVTSITS	1441
Db	1257	GLL--QTWDDARSG-----ACTGVV----CGVLRIIGSPKDFSTDRAALD-----	1296
Qy	1442	IIVSALBEDASPKASSILGWISVMGGAAGLAESAIIKGTKLATHLGFAEPDENALLKSTS	1501
Db	1297	----ALKETQNRS-----LCNDMEYSIGVCKDTNG-----KYFASKAETDNLR-K	1336
Qy	1502	ESSRIKWGTVSRDLREIVRNEEGQVIKDHSRGYTDNFMKGCEQAAILVHGDK--DGFLYHT	1559
Db	1337	ESYPLEKKCTGTDRVAAYHTHG---ADSHGDYVDEFSSSDKMLVRSKONNLEAFYLAT	1393
Qy	1560	EGNKH-----NGKGPY 1570	
Db	1394	PDGRFEALNNKGEY 1407	

RESULT 15

ABU14693
ID ABU14693 standard; protein; 1426 AA.

New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

The invention relates to an isolated nucleic acid comprising any one of the 5213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of

the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published/pct_sequences](http://wipo.int/pub/published/pct_sequences)

Sequence 1426 AA:

us-09-889-874a-23.rag

Tue Jul 6 16:41:07 2004

Db 872 DNGDLVRAISGRQ-TREYGYSAIGRLSVKTLAPDLDIRIPYATDPAGNRLDPDELHPDS 930
Qy 1131 -----VTQN-----FTYDIYCNITACHTTFADG---TEDHATFKFANPTDPCOLTEV 1174
Db 931 TLTVWPDNRIAEDAHYVYRDEYGRUTEKTRIPAGVIRTDDETT----- 975
Qy 1175 HHTHPMDPNIRLKYDKAGRINITD-NHGN-----TENFTYDTLGR----- 1215
Db 976 HHYH-----YDSQHRLVYTRIQHGEPLVESGRYLYDPLGRMAKRVWRERDLTG 1025
Qy 1216 ----LONGGSGVYGDPLNLVSKDTDLDCELYR----- 1247
Db 1026 WMSLSRPEVTWYGWDG-DRLTTVOIDTTTRIQTVYEPGSFTPLIRVETENGEREKAQRKS 1084
Qy 1248 --ETMLYNEVRNG-----EMIRLL-RTGETIIA-----QQRASKV 1279
Db 1085 LAETLQEGSGENGHVFPFAELVRLDLRLEEIRADRVSSERAWLAQQCLTVEQLARQV 1144
Qy 1280 LLTCTDSQOS-----VILTSDKQNLQSEA-YSAVCKHKSTANDASILGYN----- 1323
Db 1145 EPEYTPARKAHLVHCDHRGLPLALISEDGNTANSAEYDEWGNQNLNEENPHV--YQPYRL 1202
Qy 1324 -GERADPVSGVTHLGNVSYDPTLMEFHTPDSLSPEG-AGGINPYSYCLGDPINRSDPS 1381
Db 1203 PGQQHDEESGLY--NRHRYVDPLQGRYITQD---PMGLKGGWNLQYPL-NPQQIDPM 1256
Qy 1382 GHLSQWQWTGIGMGIAGLLLTATGGMALAAAGGIRAAIASTTTALAFGALSVTSDITS 1441
Db 1257 GLL--QTMDDARSG-----ACTGGV-----CGVLSRIIGPSKFDSTADALD----- 1296
Qy 1442 IVSGALEDASPASSILGWVSMGMGAAGLAESAIGKGTKLATHLGAPAEDEGENALLKSTS 1501
Db 1297 ----ALKETQNRS-----LCNDMEYSIGVCKDTNG-----KYFASKAETDNLK--K 1336
Qy 1502 ESSRIKWGVTRSLDREIVRNEEGQVIXHSEGYTDNFMGKGEQAILVHGDK--DGFLYHT 1559
Db 1337 ESYFLKRCPTGTDRVAAYHTHG---ADSHGDYVDEFFSSSDKNLVRSKDNLEAFYLAT 1393
Qy 1560 EGNKH---NGKGPY 1570
Db 1394 PDGRFEALNNKGEY 1407

Search completed: July 3, 2004, 05:46:37
Job time : 89 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2004, 05:43:29 ; Search time 28 Seconds
(without alignments)
3084.652 Million cell updates/sec

Title: US-09-889-874A-23

Perfect score: 8879

Sequence: 1 VVIFLKLFRITWSDNEF.....PRKILGRTEKTVFKTRP 1673

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 segs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/6CTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1638	18.4	1584	3	US-09-251-645-6
2	331.5	3.7	1377	4	US-09-711-164-467
3	317	3.6	998	4	US-09-252-991A-28424
4	313	3.5	1627	4	US-09-328-352-6604
5	306	3.4	1586	4	US-09-543-681A-5329
6	294	3.3	1439	4	US-09-543-681A-7560
7	292	3.3	1626	4	US-09-252-991A-23805
8	285.5	3.2	1596	4	US-09-328-352-5542
9	277.5	3.1	974	4	US-09-252-991A-23640
10	271.5	3.1	1665	4	US-09-543-681A-4476
11	238.5	2.7	1043	4	US-08-851-567B-61
12	235.5	2.7	804	4	US-09-328-352-5545
13	223	2.5	1128	4	US-09-252-991A-31032
14	222.5	2.5	3290	4	US-09-328-352-5486
15	220.5	2.5	1183	2	US-08-447-031A-2
16	218.5	2.5	2123	3	US-08-968-685A-10
17	214	2.4	2504	4	US-09-328-352-5821
18	212.5	2.4	954	3	US-09-251-645-12
19	211.5	2.4	10182	4	US-09-134-001C-3159
20	207	2.3	2314	4	US-09-268-347-49
21	202.5	2.3	2777	4	US-09-543-681A-6124
22	201	2.3	1739	4	US-09-540-236-3739
23	201	2.3	2057	4	US-09-489-203-2
24	198.5	2.2	1004	4	US-09-268-347-30
25	190	2.1	2385	4	US-09-543-681A-6304
26	188.5	2.1	2736	4	US-09-252-991A-30227
27	185.5	2.1	1166	4	US-09-200-650E-7

28	184.5	2.1	2142	4	US-09-540-236-3459	Sequence 3459, Ap
29	184	2.1	3696	4	US-09-134-001C-5080	Sequence 5080, Ap
30	183	2.1	1222	4	US-09-206-942-37	Sequence 37, Appl
31	183	2.1	1228	4	US-09-206-942-34	Sequence 34, Appl
32	182.5	2.1	1565	4	US-08-851-567B-59	Sequence 59, Appl
33	182	2.0	1861	2	US-09-543-681A-5434	Sequence 4, Appl
34	181.5	2.0	2315	4	US-08-790-912-4	Sequence 5434, Ap
35	181	2.0	1228	4	US-09-463-402-2	Sequence 2, Appl
36	181	2.0	1228	4	US-09-889-572-2	Sequence 2, Appl
37	180.5	2.0	930	4	US-09-134-001C-5314	Sequence 5314, Ap
38	180.5	2.0	930	4	US-09-386-962C-10	Sequence 10, Appl
39	180	2.0	1683	3	US-08-755-587-183	Sequence 183, Appl
40	179.5	2.0	1833	4	US-08-621-944A-4	Sequence 4, Appl
41	179.5	2.0	1833	4	US-08-945-567D-4	Sequence 4, Appl
42	179.5	2.0	1992	4	US-08-621-944A-3	Sequence 3, Appl
43	179.5	2.0	1992	4	US-08-945-567D-3	Sequence 3, Appl
44	178.5	2.0	1338	1	US-08-471-033-50	Sequence 50, Appl
45	178.5	2.0	1338	2	US-08-471-044-50	Sequence 50, Appl

ALIGNMENTS

RESULT 1

US-09-251-645-6

; Sequence 6, Application US/09251645

; Patent No. 6281413

; GENERAL INFORMATION:

; APPLICANT: Kramer, Vance C.

; APPLICANT: Morgan, Michael K.

; APPLICANT: Anderson, Arne R.

; APPLICANT: Hart, Hope

; APPLICANT: Warren, Gregory W.

; APPLICANT: Dunn, Martha

; APPLICANT: Chen, Jeng S.

; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS

; FILE REFERENCE: CGC1963/A

; CURRENT APPLICATION NUMBER: US/09/251,645

; CURRENT FILING DATE: 1999-02-17

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 6

; LENGTH: 1584

; TYPE: PRT

; ORGANISM: Photorhabdus luminescens

US-09-251-645-6

Query Match 18.4%; Score 1638; DB 3; Length 1584;

Best Local Similarity 31.0%; Pred. No. 9.5e-118;

Matches 493; Conservative 233; Mismatches 682; Indels 182; Gaps 54;

QY 19 EFTQANNFTSAVSGVDPRGLNIQTILGHVGN--GNLGPTLPLTSLVSPLNKTDIG 76

Db 5 DIYSNAPFGSYINTGVDPRTGQXSANIITLRENNVGNSEQT--LSLSFSLTTLNNG 62

QY 77 FGIGNFGLSVYDRKNSLSLSSTGENYK--VIEDTKVTKQKKLDNLRREKOLKENCY 133

Db 63 FGIGWRFSLTLDIKTLTFSRANGEQFKCKPLPPNNNDLSFKDKKLDKLRVYK-LDSWTF 121

QY 134 RIHKSGDIEVLGTGNNAFADLVKPKLNPAGHAIVYDMNFEATQPRNLRIYDLDGDH 193

Db 122 VYVKNKGIIELKRTGSS--DIATVALEFPDGEAFDLIYNSR----- 162

QY 194 IPLINLEYCGLIKTLTLT--FPGQEGYRTEERFL--NQLNSIHNFSLGNENPLTWSFGYT 251

Db 163 FALSEIKRVTGKTKLYKLYSGNN---CTSVYEPDDNNISAKIAFDVNNLYITVTVPYD 219

QY 252 PIKK-----NGILQWITSMATPGGLKETVYNNNQGHHPQSANLPVLVY 298

Db 220 ASGPIDSARFKMTYOTLKGVF--PVISTFRPTGTGVVELSVYKEN--GH---KVITETIYPY 273

QY 299 VTLMKQVPGACQAIQAEYSYTS--HNVYGGSGNCI--WNNKLDNLYGLMTYNGSTESRR 356

Db 274 AAALITQPGNGQPAVSKSYEYVHNFVFLGYSGRSTFSSQDNLYLVTKYKTSYIE--- 330
Qy 357 YKKEGHDQIVRTYVNNVHLLTSCCKOQNGVIOITETAYVYAIIGHNFDQSQFOLP- 415
Db 331 -RVLDGOSVSVVERVNFHLMKBAKQDNKRITTEITYNEDLSKSFQENLQOPS 389
Qy 416 KKTETWRSADNYSRSEITETTFDSGNPLTKVHKDKTKTKIISPTHWYYPAGEVDN 475
Db 390 RVLTRYTDIOQNTSREETVNIKSDDDMGNTLL-ITETSGIQK-----EYVYYPVNGECS 442
Qy 476 CPPEPYGFTFRFVKIKIQTVDSEFKDDPEKFOYRY-----SLIGSQ-----SHVTLKIEE 526
Db 443 CPADPLGFSRFLASVTKQSPDAAQVANKVHIHYIQKFTPTGAYVZYVSKYSEYIDN 502
Qy 527 RHYSAVQLNLTFOYNTDKSELGRLLKOTETCKGKNGKTSYVHKKFTYTKQDDTLQOSH 586
Db 503 KIARTFSYVNSP-----TSKSH-GSLAKITSVMNQ-----QTVTFKYEYSESEMTNA 551
Qy 587 SIITHDNFTIHRSQVESRYTGLFSDTDKDIQVQSYDKLGRLLRTRLNLSGTPYANTLT 645
Db 552 TVTGFDGAMESKNTSITVHQLKRYDVNVHITDQSYDLGRITGQIIDPSTAREIKRN 611
Qy 647 YDELANNLDQNRPPVITTTVDNGNQLNEFDGAGRHSOCLKSDG-----DKKF 698
Db 612 YVYQYPGSGDENFWP-VMEIEVDSQVRRKTHYDGMGRICSEIEQDDGANGTSGIYQCTY 670
Qy 699 YTHQOYDQEGRHHSTSYDVLNTRGOQTD-----DKVHLSMSKSYDNMGOLANTHWS 753
Db 671 KVLARQYVULGQLSKXESINDMLN--LSANPLRLATPLVTKYKXDGNGNLYSEYS 728
Qy 754 YGVSEKITVPITLTAIK-----QLOSNNSNVQKGVETTYTPSQOPIITLDEAGHL 807
Db 729 DGRILEIHDPITRTITQCVKGLGLMNIQNNF-----EQPASIKAYVPDGTI 776
Qy 808 QSCHTLTRGDWRKRETAIGQCTIYQYDNNRVIOITLDPGTIVNRKYAPFSTDLIT 867
Db 777 YSTRTRYVQGRVETTEDAEHATQIGYDVEDRIVKTLDPDGLILSAYASFSEBELIS 836
Qy 868 DIRVNGISLQOFTDGLSLTQSQDGRVWATYVAGNDQCPSTVITPDGOFIHYQOPE 927
Db 837 ALNVNGTOLGALYDGLGRVISDTVGRKTEYLYGQGDK-PIOSITPSHNKQNDLYY 895
Qy 928 LDDAVLQVANSNETQOFSVNPVTGALLKAVAGQSLTPI-VYPSGRKMWEN-INDMKMS 985
Db 896 LGSVMSKFTGTDOQNFVHSTKGTLLSA-SEGVSTQNYVFPFGVLORESFLRDNKPI 954
Qy 986 ---YLTWTLGLENGYDGLTQKTSRDTHGRVTOIKDSSIKTILNYDNLNRHIGSQVTD 1042
Db 955 SGEYLTMSGLIORHKDSFGHHVYSYDAQRLVRKTEQOAOYATFEYDNGVGLITTTTKD 1014
Qy 1043 LATGHMLTTFVDFGLNRIEGRKLCSSGHTLDIQOSWLKTOQLANRIVKLVGLORTEQ 1102
Db 1015 TTSLSQVTKIEYDAFDRKXLSLSDPSIQV-ITLSYTKNQISQISITSDGVVMKNER 1073
Qy 1103 YSVDNRNRNQKCGAECPTDKYGHISVTONFTYIYGNITACHTTFADGTEDHATFKF 1162
Db 1074 YOYDNNQRISOYCEGEQSPIDHTGRVLNQIHYHDOMGNIKRLDNTYRDKET-VDYHF 1132
Qy 1163 ANPTDPCQTEVHHTHPMDPNIRLYDKAGRVNIT-DNHCNTENFTYDGLR-----QN 1218
Db 1133 SQ-ADTQLIRI-----TSKQOQIELSYDANG-----NLTRDEKQF--LLYDQNNRLVQKD 1182
Qy 1219 GQGSV---YGYDPLNRLVSQ--KTTDLCELYYRETMVNEVRNGEMIRLRTGCTETIAQ 1273
Db 1183 RLGNLVCSQYDALNKLTAQVLANGTVNRQ-HYASGKVNTIQLGDEAITWLSDDKQIGH 1241
Qy 1274 QRA-----SKVLTGTDSQSVLTGDKQNLSCQVAYSAGKHSTANDASILGVNGERADP 1329
Db 1242 QSAKNGQSVYQYIGIDHNSVTIASQNELENALSVTFYGFPSLI---SSPLNGAQVDP 1298
Qy 1330 VSGVTHLNGYSYDPTLMRHHTPDSLSFPFGAGGINPYCYLGDPINRSDPSGHLHQAW 1389

Db 1299 VTGWFLNGYRVFNPVLMFRFHSFSDSNFPGRGGINFTYQCCDPINRDLNGLHLSAGGI 1358
Qy 1390 TIGMGIAGLLTIAT--GMAIA-----AAGGIAAIASTTTALAFGALSVTSDITSIV 1443
Db 1359 LGIVLGAIGIIVIGVLSGAGAIAGLAIAGALGALASTSALAVTATVIGLAADSIGIA 1418
Qy 1444 SGALEADSPKASSILGWSMGAGAGLAESAIGKGTKLATHLGAPEADGENALKST-SE 1502
Db 1419 SAALSEKOPKTSGLINWISAGLVLSFGISAI-----TFT-----SSLVKSARS 1463
Qy 1503 SSRIKWGTRSLDRIVRNEEGQVVKDHSR 1532
Db 1464 SQAVSAGVIGSVPLEF-----GEVASRSSR 1488

RESULT 2

US-09-711-164-467
; Sequence 467, Application US/09711164
; Patent No. 6589738
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Alllyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY
; FILE REFERENCE: ELITRA.008A
; CURRENT APPLICATION NUMBER: US/09/711,164
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 467
; LENGTH: 1377
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-711-164-467

Query Match

Best Local Similarity 3.7%; Score 331.5; DB 4; Length 1377;
Matches 288; Conservative 170; Mismatches 450; Indels 477; Gaps 75;

Qy 211 LFPQGEY-RTEFLNRLQNLNFSNGENPLTWSCFYTPGKNGILGOWITSMAP 269
Db 126 LFPGE-DCYSRSESLLVLR--GGVAKLDEGRLLAALWQ----- 160
Qy 270 GGLKETVNS-----NNWQ-----HHPOSANL--PVLXVTLMKOVPGAGQ- 310
Db 161 -ALPEELRLSHRYLATNSPOGPWLLGWCEVPEADVLPAFPVTVTLGLVDFRGT 219
Qy 311 PAIQAEYSYTHNYVGGSGNGINNNKLDNLYGLMTEYNGSTESRYKDKXGHDQIVRIE 370
Db 220 QTFHRAAGPSGEITVTDGAWRH-----FRLVL-----TTQAGR----- 255
Qy 371 RTYNNHLLTSECKQNGYIOTTETAYVYAIIGHNFDQSQFOLPKT---KTTETWESADN 427
Db 256 -----ABEAKQQ-----ALSG-----GTEPSAF--PDLPGYIYGR--DN 287
Qy 428 SYRSEITETTFDESNGNPLTKVHKDKTKTKIISPTHWYYPAGEVDNCPPEP---YGPT 484
Db 288 GIRLSAVWLTHDPE-----YF-----ENLPAAPLVRYGWT 317
Qy 485 -----RFVKIKIQTVDSEFK-----DDPEKFIQYRYSLIGSQSHVTLKIE 525
Db 318 PRGELAVVYDQSGQVASFYDDKIRGRVMAHRTGRPE--IRYD--SDGRVTEQLN 372
Qy 526 ERHYSATOLLNSTLFOYNTDKSELGRLLKQTEC--TKGNGKTSYVHKKFTYTKQDDTLQ 583
Db 373 PAGLSYT-----YQEKDRITITDLSLRREVLHTQGEAG-LKRIVVKK----- 413
Qy 584 QSHSITHDNFTIHRSOVRSRYTRGLFSDTDKDIQVQSYDKLGRLLRTRLNLSGTPYAN 643
Db 414 -----EHADGSVTQSOFDA--VGRLEAQDAAGRTTEYSPDVVTGLITRIT---TFDGR 462

QY 644 TLTYDYELNLODDNRPFFVITTTDVGNGQNLNRPDAGRHVSQCLKDS-----GDKF 698
Db 463 ASAFYNHNO-----LTSATGPDGLREYDELGRLIQETAPDGDITRYRYNPH 514
QY 699 YTHHQYDEGRHHTSYSTYLTNGRQOTDPDKVHLSMSKSYDNWGQIANHWSYGVSE 758
Db 515 SDLPATEDATGSRKMTWSRY---GQLLSFTDCSGYVTRVDHDFGOMTAVHREGLSQ 571
QY 759 KITVDPI-TLTAATKOLQSNVNVQTKVITVTSQPIQITLDFEAGHLQ-----SC 810
Db 572 YRAVDSRQGLIAVKDQGHETRYE-----YNIAGDLTAVIAPDQSR 612
QY 811 HLTTRDGDWRVREKTDALGCTQITVQYDNNRVIQITLPDGTIVNRKYAPFSTDLITDIR 870
Db 613 NGTOVDANGKAVRTTQ--GLTSEMYDAAGRVILTSNGS-----HTTFYDVLRLIQ 666
QY 871 VNGISLGQOTF--DGLSRKLTQSQGGRWAYVYSAGNDQCPSTVITPDGFIHYQYQPEL 928
Db 667 ETGFDGRTQRYHDLTGRLIRSEDEGLV-----THWYD-EA 702
QY 929 DDAVLQVNASNEITQFSSNPVTGAL--LKAVAEQSLTPIY---PSGRL--KMNENDM 981
Db 703 DRLTHRTVKGTAEARWQDE--RGWLTDSHISEGHRVA--VHYRYDEKRLTGERQTVHP 760
QY 982 KMSYLM-----TLRGLNGYTDLTGTIQTISRDTHGRVTOIKDSSIKTILNYDDL 1033
Db 761 QTEALLWCHETRHAYNAQGLAN--RCIPDSLPAVEWLTG-----SGYLAKMLGD-- 809
QY 1034 RHIGSQVTDLATGHMLTTFEF--DGLNREIGRKLCDSSGHTLDIQSWLKQTOANRIV 1091
Db 810 -----TFLVEYTRDLRHRETLR-----SFGVELTATYTPAGLOSOQ-- 846
QY 1092 KLNGVLTQEQYSDSRNLNOYKCDGAECTDKYGHISIVTONFTYDIYGNITACHTTFA 1151
Db 847 HLNLSLS--DRDYTWNDNGELIR-----ISSPRO-----TRSYSTTGRLTGVHTAA 893
QY 1152 DGTEDHATFKFANDTDFC--QLTEVHTHPD-----MPDN-----IRLYDKAGRVIN 1197
Db 894 -----NLDIRIPYATDPAGNRLPD--PELHPDSTLSMWPDNRIARDAHYLYRYDRHGLTE 947
QY 1198 ITD-----NHGNTENTYDTLGLRONGQ-----GSVGYDPLNRLVSKTDT 1239
Db 948 KTDLIPSGVIRTDDERTHRHYDSQHLVHYRTQVEEPLVESRYLYDPGRVAKRVR 1007
QY 1240 LDCEL-----YY-----RETLVNE-----VRNGEMIRLR-----TGETI 1270
Db 1008 RERDLTGWMSLSRKPQVTWYGDGDRLTITQNDRTRIQTIVQPGSFPLIRVETATGELA 1067
QY 1271 IAOORA-----SKULLTGTDSQQSVIL-----TS 1294
Db 1068 KTORSLADALQSGGEGDGGVPPVVLVQMLDRLESEILADRVSPESRRWLASCGLTVE 1127
QY 1295 DKQV-----LSQEA-----YSAYGKHKSTAND--AS 1318
Db 1128 QMOMQMDPVVTPARKIHLXCHDRGLPLALISKEGTENCAEYDEWGNLLNEENPHLOQ 1187
QY 1319 ILGNGERADPVSGVTHLNGYASYPDPLMRFTPDLSLSPFG--AGINPYSYCLGPPINR 1377
Db 1188 LIRLPQQYDEBSGLAY--NRHRYDPLQGRYITQD---PIGLKGMNFYQYPL-NPVTN 1241
QY 1378 SDPSG 1382
Db 1242 TDLG 1246

RESULT 3

US-09-252-991A-28424
; Sequence 28424, Appication US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28424
; LENGTH: 998
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28424

Query Match 3.6%; Score 317; DB 4; Length 998;
Best Local Similarity 22.0%; Pred. No. 2.2e-15;

Matches 222; Conservative 129; Mismatches 362; Indels 296; Gaps 47;
QY 504 EKFIQRYSLIGSQSHVTLKIERHYSATQLNLSILFQYNTOKSELGRLLKOTECTKGN 563
Db 112 ERAWSYSYNALG-----LIBRADGPRTDVQDVTLYAYD-----SRGNL 149
QY 564 GKTSYVVKFTYTKQDDTLQQSHSITTHDNFTIHSQVRSRYTG--RLFSDDTDDKDIVQ 621
Db 150 TQVTNALGOVTRLGDYDERGKPGSITDANGVTSSLA-----YTGVDGMLASVSTAGSTR 204
QY 622 MSYDKLGRLLTTLASGTPYANTLYDYELNNLQDNRPPFVITTTDVNGNQLRNEFDGA 681
Db 205 PDYDAVGQTRVTRGEGD---SWLSYEV-----DDARR--LVAIGNLGERLEYDVTK 252
QY 682 GRHVSQCLKSDGDKGFYTHFQQ--YDBQGR-----HHTSYS--DYLTNGRQOQDP 730
Db 253 GNRTAQRIKASGS-----LVROQWAYDELGLRLRAVGAGGQTRSFAYDLNDNPVGETNP 308
QY 731 DKVHLSMSKSYDNWGQIANHWSYGVSEKITVDP--TLTATKOLQSNVNVQTKREVITYT 790
Db 309 RC--FAHSQAFDALDLVQSDPLGKTRLAYD-----AODNLTEVKDPRGVTRY 357
QY 791 PSQQPIQITLPEAGHL-----QSCHTLFDGHDVRKKTDAICQCTIYQYDNNRVI 843
Db 358 E-----YDGLNGLRILVSPDSGTTTFHDAAGNVIRRTDAEGAVTEYRYDALNRLV 408
QY 844 QITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQOTFGLSRLTQSQDQGRVWAYTSA 903
Db 409 ERRSP-----SDPSLDVQYRYDLTADGNGIGRLGAIGARDS-----446
QY 904 GNDQCFSTVITPDGQFIHYQYQPELDDAVLQVANSNEITQFSSNPVTVTGALLKAVAEQSL 963
Db 447 -----LVRYD--ERGNLVEQVRSIRLQDQTLDRVTVRYDAA-----NQL 485
QY 964 TPIYVPSGRLKMNENDMKMSYLTWLRGLENGYTDLTGTIQTISRDTHGRVTOIK--- 1019
Db 486 LEIGYPS-----GLAIGY-----PRNAGGQVASTLAVG 514
QY 1020 DSSIHTT-----LNYDDLNRHISQVTDLATGMLTTFVFDG-----LNREIGRKLCD 1069
Db 515 DKAPSTLVQIAYLPFGPLQR-----LTWNGITLSREYDQDYQLLRKVG-----560
QY 1070 SGHTLDIQSWLKQTOANRIVKLVGRVOR-----TEQYSYDSRNLNOYK--CDGAE 1121
Db 561 -----PWQSDYQ-----HDANGNIQQRHSLWGLTDYQYDPLDLTEERGQVGG-- 604
QY 1122 PTDKYGHISIVTONFTYDIYGNIT--ACHTTFADGTEDHATFKFANPTDPCQLTEVHTHPD 1180
Db 605 -----RSYADVAGNRTQSDNPASGGTASSQDYQA-----636
QY 1181 MPDNRL-----KYDKAGRVINITDNGHTENFTYDILGRON-----GQSVTVGYD 1227
Db 637 -PDSNRLTAIGAQAVTSDAAG---NLQDRA--ARKLAYDAQGRQLQSVLSLQDQQAERYN 691
QY 1228 PL-NRLVSKQTDTLDCELIYVRETMVNEVRNGEMIRLIRTG-----ETIIAQORA 1276
Db 692 ALGRIVKLTESITTYLYGPDGQLLGEAEHDSGRKLRAQYLYLWLSPLPLATTDADYDA 751

QY 1277 S-----KVLITGTSQSQSVILTSK--QNLQSEAYSAYGKHSTANDAS--ILYNGER 1326
 Db 752 QGKVGNTLLYLHGDHLDTPRLATDASGQIAWQWSDAFGRGEALSQSGSTQVNLRFPGQY 811
 QY 1327 ADPVSGVTHLNGCYRSYDYLRFHTPDLSLSPG-AGGINPYSYCLGPPINRSPSG--- 1382
 Db 812 YDAESGLHY--NYFRDYDPETGRYVESD---PIGUSGGVNTYGVQGAFLNRIDPLGLAA 866
 QY 1383 ---HLSQWAWTIGMGI---AGLLTIATGMAIAAGGIAAIAASTST 1425
 Db 867 IEIDIPKSAVDWTPGNIRLPACRLD---GGVLLVASISGATPQADSST 911

RESULT 4

US-09-328-352-6604
 ; Sequence 6604, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 6604
 ; LENGTH: 1627
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-6604

Query Match 3.5%; Score 313; DB 4; Length 1627;
 Best Local Similarity 20.4%; Pred. No. 1.1e-14;
 Matches 303; Conservative 177; Mismatches 506; Indels 498; Gaps 73;
 QY 75 IGGIGFNGSLSVYRKNSLSLSTGENY--KVIETDKTV-----KIQ 115
 Db 369 IFPMATEFTTHVDAQGLIQQDSIRTVSNLVQMDAIFGARWVPTTKISRKFKYT 428
 QY 116 QKLDNLREPKLKNCRVRIHKSGDIEVLTFGNNAFDLKVKKLNPAGHAIY----ID 172
 Db 429 SKKKCHKYVLENGLEVIC-----LDGRAIDLPLKK-----GQSIYDPIEQ 468
 QY 173 WNFETQPLNRI-----YDDLGDHDIPLLNLEYQGLIKTILTFPQKQEGYRTEAR 224
 Db 469 YTYTVLSQGLHLIAYGEDEKRYEYKGYEDYRLSYIE-----RKNGFKVALR 514
 QY 225 F-----LNRQLNSIHNFSIGNENPLTWSFGVTPIG-----KNGILGQWITSMTAPGG 271
 Db 515 YDHVSTDNKTIISLILFKQDDNLLAHLALQLTPOGLVSDIWTIKNGQLDRVLAS----- 568
 QY 272 LKETVYNSNNQGHFPPQSANLPLVYTLMKQVPGAGQPAIQAEYSVTSH-----NYV 325
 Db 569 -----YDYOQGG-----DLVQATNEPFAASYQYTHLITRYDILT 604
 QY 326 GGGNGIWMNKLDNLNLYGALTEYNGVSTBSRYKQEGHDQIVRIERTYNNHLLTSECKQ 385
 Db 605 HRGMNLKWDGILPTSKAIEWADNASRASKLEWDK-----NIRKT-----TVLDVEGNS 653
 QY 386 QNGYIQTTEYAIIGHNFDSPQSPQFOLPKTKTETWESADNSYRSIEITETPDESQNP 445
 Db 654 THEYVIDIGYTVRIYVDPNFE---ECFPRDDAKNITLHIAKDGSK---TSYTVDERGNVL 707
 QY 446 TKVIKDKTKQKISPSSTHWEYYP-----AGEVDNCPPEPY-----GTFRFVKIITQPYDSE 498
 Db 708 TTTQDDGAT-----SYFEYDEKNQLTGMVDAEQGRWFKQYDGSNGNLKEI----- 752
 QY 499 FKDDPEKFTQYRSLIGSQSVHLLKIEERHYSATOLLNLTFLQYNTDKS-----ELGRLLK 554
 Db 753 ---DPLK-----HETAVY-----YVAGLVTSITDAKGGSKSLKYDDQGNLIS 792
 QY 555 QTEC-----TK-----GENGKTYSVH-----KFTYTKQDDTLQO----- 584

Db 793 YTDCSGKETKQYDERGVSIVISINALNQKVBYFYFTELTLENREPIIKGLPLNAFGQLEKI 852
 QY 585 SHSITTHNFTHRSOVSRYTGRLSFSDTDTKDVTCMSYDKLGRLLTRT--LNSGTPYA 642
 Db 853 KHAQGTBEHF-IHRE-----GRLLAHVDPKQNIYREYDEAGLLISRTDALN----- 899
 QY 643 NTLTYDYE-----LNNLQDNRPFPVITTTDVNGNQLNEFEDGAGRHYVSQCLKSDSDGKRF 698
 Db 900 HKLKYKWRDLRLRLTLINE-----NGASYQFFYDVASRLV-----KEIDFDGKE 943
 QY 699 YTIHQYQYDEO-GRHTS--TYSDYLTNGROOTDP-DKVHLSMSKSYDNMGQJANTHWSY 754
 Db 944 TYVH---YDEKSGQLATSIEVASYGQDLKDRAPKDRIOQFI---FDSMGRLEQRTAGY 997
 QY 755 G-----VSEKLTVDPI--TLTATKQLOSNNVQTKKEVTTYTFPSQOPIQITLFDSEA 804
 Db 998 GHYGLEEEKQTEFAVDYMGRIIQAKNAOSNLQ-----WFDAA 1037
 QY 805 GHLQSCHTLTRDGDWRVRKETDAIGQCTI--YOYDYNRVIOITLPDGTIVNEKYAPFST 862
 Db 1038 GNLVOEH-----QQDYKINKTAVWKHQDEINDEIKTRPDGQVID----- 1078
 QY 863 DTLITDIRVNGISLQOQTFDGLSKLTQSDGGRVWYATYSAGNDQCPSTVITPDGQFIHY 922
 Db 1079 -----W-LTYGSGHVQ-----SLIVNGQDFVSF 1100
 QY 923 QYQPELDDAVLQVASNETIQOFSYNPVTGALLKAVAGQSILTIPIYPSGLKVENINDMK 982
 Db 1101 -----ERDDLHREIA-----RHYANGVSQEQOYDLAQLKSKQM----- 1134
 QY 983 KMSYLTWLAGLENGYTD-----LTGTICKISR-----DTHGRVTOIKDSSI-KTLLNY 1029
 Db 1135 -----LSEHENGYQNYKRRHNNALQTSQVQLVQYDKTGELTAIRDTREGNIAYKY 1187
 QY 1030 DDLNRHIGSQVTDLATGMLTTFVEFDGLNREIGRKLCDSSGHTLDIQOSWLKTOOLA-- 1087
 Db 1188 DPGRELLIE-----ASSKLKETFSFDPASNIL-----DSYHSQVQSHSQKLDTSYG 1235
 QY 1088 -NRIVKLVGLQR--TSQYSYDSNRNLNRYKQDGAEGFTDKYGHSIVTONFTYDIYGNIT 1144
 Db 1236 YNRLV--NNVVKEYLDQOYQYDAYGQIROKTSQGL-----NLEMDVYGRMV 1281
 QY 1145 ACHTTFADGTEHATFANPTDPCQLTEVHTHTHPDPMNIRLKYDKAGRVINITDNHG 1204
 Db 1282 KSRNS-----QYTAHY-----RYDALGRRIQKWSKHH 1309
 QY 1205 TENFTYTLGRLQGSVYGYDPLNRLVSKTDTDLCELYYEETMLVNEVRNGEM--IR 1262
 Db 1310 T-----GOEQN---IITYGWDG-DT'LAYESTEELTKHYIYEKDSFVPMLOAVLSPIE 1357
 QY 1263 LLRTGETI-----IAQGRASKVLLTGTDSQSVLTSKONLSQE-----AYS 1305
 Db 1358 LHQTPDSDRPYNHRDPLWKTEKEGEPDDWFWFCHDLGTQEMWTDHTGAILWKAERY 1417
 QY 1306 AYGHKXSTANDASI-----LYNGERAPVSGVTHLNGYKSYDPTLMRHTPD 1355
 Db 1418 AWGECKAEEAKSNPFENSEIISNNIRFQGYQFDEETGLHY--NRYRYYSFYGRFVSKD- 1474
 QY 1356 LSPFG-AGGINPYSYCLDDPINRSDPSCHLSWQAWTGIGMGIAQ 1398
 Db 1475 --PIGLGGNNVYVA-KNPIWTWIDSKGLSTLNNLGL-GVKG 1514

RESULT 5

US-09-543-681A-5329
 ; Sequence 5329, Application US/09543681A
 ; Patent No. 6605709
 ; GENERAL INFORMATION:
 ; APPLICANT: GARY BRETON
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A

; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1993-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 5329

; LENGTH: 1586

; TYPE: PR

; ORGANISM: Proteus mirabilis

; FEATURE:

; NAME/KEY: UNSURE

; LOCATION: (1576)

; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.

US-09-543-681A-5329

Query Match 3.4%; Score 306; DB 4; Length 1586;

Best Local Similarity 22.1%; Pred. No. 3.6e-14;

Matches 243; Conservative 129; Mismatches 385; Indels 342; Gaps 59;

QY 499 FKDDPEK-----IQYYS-----LIGSQSHVTLKIEERHY-SATQLLNS- 537

DB 486 FEANPHKINTLRMKSADEHNALHYRYANDGELVQIHDDAVLTDIRLHYDEITQRCSV 545

QY 538 -----TLFOYNDKSELGRLLKQTECKTGKNGKTYVVKFTYTKQDDTL---QQS 585

DB 546 TRHQOEKTLVITYDAQ--RLQVVTNADK-----RVTRBFGWDDSEGLMAMHQYA 596

QY 586 HSITH-----DNFTIHRQVRSRYTGRFLSDTDTKDVTQMSYDKLGRLLTRTLNSGT 639

DB 597 TVSSHYRQWRFDAFTIEDNEPEWVVEHMLKDGKRCLEHTELYDLAQRLLTVETGGE 656

QY 640 PYANTLTYDELNLQDNRPFFVITTTDVNGNQLRNEFDGAGRHVSQCLKDS---PDGG 696

DB 657 -----TTFRRNEQQIIEYTNALNETWPEWD-----TSRLKKAIPDGE 699

QY 697 KFYTIHQOYDEQSGRHTSTYDYLNGRQOTDPDKVHLSKMSKYDWMGQ-----IANT- 750

DB 700 WGYT-----YBERG-----NLQWTDQEQ-----OSTCYD-WKDDFAFPTAOTL 737

QY 751 -----HWSYGVSEKI--TVDP-----ITLTAIKOLQSNNNVQTKGVITYTTPSQOPIQT 799

DB 738 PNGAAWHWEYNEHGDIRVIDPLGHITRLAWDDQGLGQVDAKGNETHRYNARGOLIE 797

QY 800 LFDGAGHLQSCHTLTROGWRVREKETAIGCTIYQVDYNNRVLIQITLPGTIVNRKYP 859

DB 798 QRCDSGYPT-----TUTYDDWGQLRLTNAQNETTTTTFSEAGLLTECLPBDT--ENRDIY 852

QY 860 FSDTL--ITDIRVNGISL--GQ-----QTFDGLSRLTQSQDQGRV 896

DB 853 DATQGLVITDAGERHILLRNRGQVIAREDPAGHMLHFHYDTFGMQALENEQ--GEQ 910

QY 897 WAYTYSAGNDCPSVTITPD--GQFIHYQY-----PELDDAVLQVASNE 939

DB 911 YRFSYDALHR-----LTDEHDLIGQOKHYQYDVMGNVTQIKTTPGPSIDTPI--PLSPQ 962

QY 940 ITQOFSYNPVTGALLKAVAEQSLTPIYPSGRLLKMNINMKMSYLM-----TLRGLE 994

DB 963 VT-TFGYDKV-----GRLLFRENADY-RTEVLYOPLSVTLRRVP 999

QY 995 NGY---TDLTGTQKISRD-----THGRVTQ--IKDSSIKTTL--NYDDLNR----- 1034

DB 1000 MAVWHEAERTGTARVEYQDALFTYDKVGQLVREASARGDYQHYYDVLGNITRTELPHQ 1059

QY 1035 -----HIGS---QVTDLATGMLTITVEF--DGLAREIGRKLCDSGHTLDTQQSWLKT 1083

DB 1060 RAEFLYGGSHLOOTWRNEQUTVLAIEYQRLRHRETLR-----TSG-ALDNETGYDCR 1114

QY 1084 QQLANRIVK-----LNGVLQRTQEQSYDSNRNL-----NOVK 1115

DB 1115 GRITHQVARQMASQFVTPVIDR--RYWDBKQNLIIERSVSYGQTEVFTAGHWYHSQ 1172

QY 1116 CDGACETDXYGHSIVONTYIYINITACHITTFADGTEDHATFKFANPTDPCQTEVH 1175

DB 1173 YDPLGQLTAHLG-SVQTEHFLYDAANLL-----TRPH-----TEAP 1208

QY 1176 HTHPDMPDNIRLKYDKAGRVINITDN-----HGNTEN-----PT 1209

DB 1209 HNVQSGSKFDYRDYDGFGRMVSVYKSGSSGQRYHYDSHRIIAVDIDQGLGQYQAEYR 1268

QY 1210 YDTLGR-----QNGQSVYGYDPLN-RLVSQKTDITLDCBLYYRETW 1250

DB 1269 YDILGRRIEKRLWKASAIANTVTYHQHEPDDEVYTFGVMGRVLVSEHSSAAAPHHTTVVH-- 1325

QY 1251 LVNEVRNGEMIRLLTGETIIAQORASKVLLTGTDSQQSVILTSQKQLNSQEAYSVAG-- 1308

DB 1326 AYNDQSVYTLARIECT-DNPLNFORALYTHSHLSGLPEALTNSEGEIWMQOQYSAWHL 1384

QY 1309 -----KHKSTANDASILGYNGERADPVSVTHLNGYRSYDPTLMRPHTPDLSLSPFG-AGG 1363

DB 1385 QRCRTPTSTFNRQNLRFQOYFDKETGLHY--NTRYIYAPDLGRFTQOD---PIGLAGG 1439

QY 1364 INPVSYCLGDPINRSDPSG 1382

DB 1440 INLYAYA-ENPLTWDPWG 1457

RESULT 6

US-09-543-681A-7560

; Sequence 7560, Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCES: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A

; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 7560

; LENGTH: 1439

; TYPE: PR

; ORGANISM: Proteus mirabilis

US-09-543-681A-7560

Query Match 3.3%; Score 294; DB 4; Length 1439;

Best Local Similarity 20.3%; Pred. No. 2.6e-13;

Matches 251; Conservative 141; Mismatches 386; Indels 460; Gaps 62;

QY 347 YNKGSTESRYK-----DKEGHDQIVRIERTYNNYHL-----LTSECKQONGYIQ 391

DB 376 YHYPFPDDKGYSRSLSHVDNVEQORFH---YNEHHQLIHITGCDLNECEYQS--PQ 429

QY 392 TTE-----TAYYAIIGHNFDSPSQF-----QLPKTKTET-----WR 423

DB 430 LAEKTVSRLTAYQVNPQHIRRLCAYFNESAQLIRVEQQTNHPYRQFGTWAGVMAWH 489

QY 424 SADNSYRSEITTTTDESQNLTKVTKKTKQKILSPSTHWEYYPAGEVNCPEPYGF 483

DB 490 SKYGLRSEYRWALSEDN---LWRVIENTKTS-----GESYRLEYDDINL 531

QY 484 TRFVKKIITQPYDSEFKDDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQLL----- 535

DB 532 TR-----TAY-----WHQGSTFWQLNH-----DHQIHIDRTGKITALIWDFGLPC 575

QY 536 ---NETLFQYNTDKSELGRLLKQTECKTGKNGKTYVVKFTYTKQDPTLQOSSHITTHD 592

DB 576 GCRNALGHTH-SEWDALGRLLSITD---CNGNQI-----RWQYQNERERL-----IT--- 619

QY 593 NPTIHRQVRSRYRGLRFSDDTKDIVTQMSYDKLGRLLITRLNSGTPYANTLYDIYELN 652

DB 620 -----VFPDNT-----SRLAYSLGRLIKEI-----SLPHQITTEYRDFK 656

QY 653 NLQDDNRPFFVITTTDVNGNQLRNEFDGAGRHVSQCLKSDGDKFYTHIQQYDEQGRH 712

657 TTL---RP-----TIRIDAKQSRSEFLWNR--GOLLRHTDCSGK---QHIMCYDDEGRV 703
QY 713 HTSTYSYLTNGRQOQTDPKVHLSMSKSYDNGQIANTHWSYGVSEKITVPIITLTKQ 772
Db 704 VSO-----FNALQE-----ATEYQYDEVGHARI-----ILPD-----NSTVQ 736
QY 773 LQSNNSNVQKVTYTPFSQPIQITLFEAGHLQSCHTLPRDQMDVRKKTDAIGQCT 832
Db 737 LAMNAAGLLTHHQKNDNTPCQWY-----NAFGRVTEIDKLARI 777
QY 833 IYQYDNNRVIOITLPG--TIVNRKYAPSTDTLITDIRVNGISLGQOTPDGLSRITQS 890
Db 778 HYHNAAGALISNANGRYLNRD-----ASRLVEELRDE--TLLOTTYNAGRLVEE 832
QY 891 QD--GGRVWA-----YYSAGNDQCPSTVITPDQCFIHYQYQPELDDAVLQVASNEITQ 942
Db 833 AHLGDRVTSAPRTILLDYDAAGNLVKRETLD-----RYQYQWDSMNRL--VASKQBNQ 886
QY 943 -----QFSYNPVTGALLKAVAGOSLPIYPS 970
Db 887 RLEMLGQANQVHTVDALRIIREQTDGDIVEFAYNNELNLSRLTLPQGSANLWYGS 946
QY 971 GRKLMENINDMKMSYLTWLRGLNGYDITLCTIQKISRDTGRVTOIKDSIKITLLNYD 1030
Db 947 GHATAIN-----HLVDSR-----SOLITEPERD 969
QY 1031 DIARHIGSQVTDLATGCHMLTTTIVEPDLNREIGRKLCDSSGHTLD---IQOSWLKTCOLA 1087
Db 970 DLHREIS-----RTQCELTQYQYDKUGRTISTFSSRDQKPLNGITLWRKWFYDQ-- 1021
QY 1088 NRIVKLVGLQRTQYQYSDSRNLN-----QYKCDGACPT 1123
Db 1022 GNLCAMEDTVRGVVEYLYDSQRLKVASSENLDAMLFYDRADNILLRPPQSEMDAHSP 1081
QY 1124 -----DKYGHISIVTQNTYDIYGNITACHTTFADGTEDHATKFAKNTDPCQLTEVHH 1176
Db 1082 LEISPOGDKL--ROFQWHYQYDAYGNVIA-----RKYRN-----OSSQYI-- 1120
QY 1177 THPDMPTNRLKYDKAGRVINITDNGNTNFNTYDTLGR-----RKYRN-----OSSQYI-- 1120
Db 1121 -----AYDGNRLV--IAHQGIKAQVHYDALGRRIHKTVENRESQVAKRQETHF 1168
QY 1216 -----LQNCQGSVY-----GYDLNRLVSKTDTLCELYVRETMVN-- 1253
Db 1169 IWQGLLEQDINTGKHQTCYEBHGSYTPPLAVIVKSSGF-----HYWHHCDSINAPL 1224
QY 1254 EVRNGEMIELLRTGETIIAQ-----RASKVLLTGTDSQOSVILTSDKQNLQSEAYSA 1306
Db 1225 EVTNAQ-----GNTWISGKVERFGRVSSPLSPYSSPER--VMASEFQNLA----- 1268
QY 1307 YGKHKSTANDASILGNGSRADPVSGVTHLNGVSYDPTLMRFHTPDSLSPFG--AGGIN 1365
Db 1269 -----YAGQYFNETGL--HF--NTFRYDQIGRIMPD-----PIGLLGIN 1308
QY 1366 PYSYCLGDPINRSDPSG-----HLNQWATGI 1392
Db 1309 LYQYA--PNPLGMDPWLSSQEMVRVHRTSVLEGLEGI 1345

RESULT 7

US-09-252-991A-23805
; Sequence 23805, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196, 136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23805
; LENGTH: 1626
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-23805

Query Match 3.3%; Score 292; DB 4; Length 1626;
Best Local Similarity 21.1%; Pred No. 4.6e-13;
Matches 204; Conservative 125; Mismatches 316; Indels 320; Gaps 46;
QY 597 HRSQVRSRY---TQRLPSDSTDIDIVTOMSYDKLGRLLTRTLNSGTPY----- 641
Db 633 HTQVLAHYRCNRCQQLIEAGNALYEAERYDARNVILERQLAGGASFFEWEGEKQA 692
QY 642 -----ANTLTVD--YELNN-----LQDNRPPFVITTTDVGNGOLRNE 677
Db 693 RAVHWASFPQWDSRYVWNEGDSVTAINAGSEEVYVDNAR--LVYQVDPDGETLRE 750
QY 678 FDGAGRHSVQCLKSDGSGKPYTHHQYDQGR-----HHTSYSS----- 718
Db 751 YDEKQQLVAE--RDPLG-----TITEYRYDQAGLEALPAEGEPTCYSYFDGFRVSR 803
QY 719 -----DYLTNGR---QOTDPD--KVHLSMSKSYDNGQIANTHWSYGVSEKITVPIITL 767
Db 804 GEAQWKYERNAGDITROTDEGNVTHY---AYDRGCLVEITHADGSLHQLTWPLQG 859
QY 768 TATKQLOSNNSNVQKVTYTPFSQPIQITLFEAGHLQSCHTLTRDQMDVRKETA 827
Db 860 LIEEQLPDGS-----VRRYRYDTLGRQITRODESGAI----- 891
QY 828 IQOCTIYQYDNNRVIOITLPGDIIVNRKYAPSTDTLITDIRVNGISLGQOTPDGLSR 887
Db 892 ---TRFQWDAAGRLSQITLPG--ASRY-----RYN--AYGKVT----- 924
QY 888 TQSQGGRVWATYTSAGNDQCPSTVITPDQCFIHYQYQPELDDAVLQVASNEITQFSYN 947
Db 925 SECDCQGRMTYEV--LDLHLVSRINPDPGSQLRYR--ENARLLLSIENENERGERYLD 981
QY 948 PVTGALLKAVAGOSLTPYVPSGRLKMNENMDKMSYLTWLRG--LE-----NGVTDL 1000
Db 982 -----YHNGLSIQCETGCFDGRRTAYRYDLKQLEKTEYCDGSEL 1022
QY 1001 TGTIQKISR-----DTHGRVTOIKDSIKITLNYDNNRHIGSQVTD 1042
Db 1023 RTTYQDSTGRLAKTLAPDGNRVYRYDTLGRVAVDDGTWPLAYEYDLRDR----- 1074
QY 1043 LATGHMLTTVE--FDGLNREIGRKLCD-----SSGHTL---DIQOSWLKTCOLA 1089
Db 1075 LVREHQWATLHYAYDALGQLIHCRLPDGNRVYRYQGTGTLASIDLNGSLTRHQFGSG 1134
QY 1090 IVKLVGLVQRTQYQYSDSRNLNOKYKCGACPTDKYGHISIVTQNTYDIYGNITACHT 1149
Db 1135 RERQRCQGLLSQYHYDEQGLLAHQVSQK-----RHLYQRYRYDASGNLAA----- 1183
QY 1150 FADGTEDHATKFAKNTDPCQLTEVHTHHPDMPNIRLKYDKAGRVIN----- 1197
Db 1184 IEDSRKGIKRSFHY--DPLD--RLJGVGRGETE--SFVHDPAGNLLAQCGQFARQMEV 1235
QY 1198 ---IT-----DNHGN-----TENTYDTLGR-----LQNGQSVYGYD 1227
Db 1236 RGNLLTQGBRHFYDAHNLVRERRGTGOKLVTEYSYDQCHRLIGVSPDGGQVYRYD 1295
QY 1228 PLNRLVSKOTDLDCELYYRETMVLNVRNGEMIRLRTGETIIAQ--ORASKVLLTGT 1285
Db 1296 AFGRIAKQVDG-----RNTFPLWL--GERLLAESGRHRYKTYLYEED 1336
QY 1286 SQOSVILTS-----DKQNLSCB-----AYSAYGK--HKSTANDA 1317
Db 1337 SFRPLALDGGEGPEQVPCYVQDHLGTPOELTRABGLCWSARYRAYGNLKLDAEVD 1396
QY 1318 SILGNGERADPVSGVTHLNGVRSYDPTLMRFHTPDSLSPFGAGGINPYSYCLGDPINR 1377

Db 1397 NPLRFGQYDAETGLHY--NHRYYNFSTGRYLPDPRL--AGGLNSYRY-VENPTGW 1451
QY 1378 SDPSG 1382
Db 1452 VDPLG 1456

RESULT 8
US-09-328-352-5542
; Sequence 5542, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Berton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-032A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5542
; LENGTH: 1596
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5542

Query Match 3.2%; Score 285.5; DB 4; Length 1596;
Best Local Similarity 18.1%; Pred. No. 1.4e-12;
Matches 202; Conservative 179; Mismatches 387; Indels 349; Gaps 49;

QY 440 ESGNPLTKVIXDKTKIISPTWYPPAGEVDNCPPEYGTFRVKKIOTPYDSEF 499
Db 337 QSGKSIYSIGAERVQ-----HADFYLP-----KIGFS-PIRQ-----YNSQM 373

QY 500 KDDPEKTIQVRYSL-----IGSQSHVTLKIEERHYSATOLLNSTLFOYNTDKSELGRL 552
Db 374 DEFQSVGARGWMPFNMTOQNAQGLFIDSKGRKH---QLPVSIIFFEYVEPYE-GWI 429

QY 553 LK-----QTECTKGNGKTYSVVHKFTYTKQDDTLQOSHSHITTH-----591
Db 430 IKPLKNGELILDFGGEWRSHFQSPDGKGYLVKMKNETSQEBELLE-YLLDHLIAVLKV 488

QY 592 DNFTIHSSQVRSRYTGRLFLSDTDKDIVTQMSYDKLGRLLTRINSCTPVANTLTVDYEL 651
Db 489 INFKLKQAEVELKF-----AFNEQKIIIAVFL-----515

QY 652 NNLOQDNRPPFVITTTDVNGNQLRNEPFGAGRHVSQCLKSDGDKGFYTHITQYDEQGR 711
Db 516 ----DDKAEPLA-----RYEYDTQG-----NLKALDQNG-----HRTYBYNQF 551

QY 712 HHTSTYSDYLTNGRQOTDPPKVLHSMKSYDNNGQIANTHWSYGVSEKITVDITLITATK 771
Db 552 HQUTRYTD--RTGRGQNIYESTEAKAKAIEEWADDGSEH-----TKLKWHP-----RLR 599

QY 772 QLOQSNMNVQTKREVITYPESQPIQITLFDAGHLQSCHT-----LTRGDWRVRKE 824
Db 600 QVA-----VYDAVDVPIY-----YFDDLGFTYRKLADGRSWSRDKGRITQK 645

QY 825 TDAIGQCTIYQYDNYNRVIQITLPDGTIVNRKYAPFTDITLDIRVNGISLQGFQDGL 884
Db 646 IDFDGRETOQYNDQDLVKIVQPNNGGIIRFAY-----NKK 681

QY 885 SRLTQSDQ--GGRWAWYTSAGNQCPSVTITPDGOFIHYQYQ-----PELDDA-----VLQV 935
Db 682 GNLVEIKDPGSLWKREYDENRN--VSKEINPUGHITQYKNNNDNOLVEIDAKGVKKI 739

QY 936 ASNEITQQPSYNPVTGALLKAVAGQSLTPIYFSGRLKVENINDMKMSYIWLTLRGLN 995
Db 740 QYNELQMSIYTD-----CSGKSTWEYDEDCALTAEQTANNKVQYFYFSGRDK 790

QY 996 GYT-----DLTGIIQKISRTHGRVTQIKDSSIKTTLNTVDDLRH-----1035
Db 791 GLOQSIYDPGLKEYFEHDEGRLLK-HTDTKGLVTEYKYNQVGLLEQRIDANRHSVAYQ 849

QY 1036 -----IGSQVTDLATGH-----MLTTTVEFDGLNREIGRKLCDSSGHTLIDIOQSWL 1081
Db 850 WDKQGRITKLINQQAELFGYNPYGLIREQAFDQBEKHYS---YNGRLFIQIRRPNI 906

QY 1082 KTO-----QLANR-IVKLVNGVLTQTEQSYSDSRNRLNQYKCDGACBPTDKYGHSI---1130
Db 907 LTOFDYADGQIASKSFTHLTQKQTEQFDYNLNSQLSRASNEVSEQ--IDLYRNALGQL 964

QY 1131 -----VTQNFYTDIYGNJ-----TACHTTPADG-----TEDH 1157
Db 965 VREHQHYKIPELKPLTAVLHVEYDELGNLIKIRPDCGHTLNHLVYGGSHIYAIGLNNQEV 1024

QY 1158 ATF-----KFNPTDFCQLTEVHHHTHPTDMPDNIR 1186
Db 1025 VSPORDDLHRETRILLANGLMOTKQYNDVGLLSQFNQPEQETQDVLQYQAHN-----K 1078

QY 1187 LKYDKAGRVINITDNHNTENFTYDGLGRNQOG-----SVYGYDPLNRLVQKTDLTLC 1242
Db 1079 YHYDKVYLLSQVEDSRGLKLYQYDDIGRLIAAQSLLHKTESFNFDPAGNLIDSESVLSPA 1138

QY 1243 ELYYRETMVNEVRNGEMIRLLRTGETIIAQORASKVLLTGTDSQSVILTSDKQNLQOE 1302
Db 1139 QI-----KNNLIKSYKGKHQYQVDQGNVTEIIQ--AGKNLKLTDWQNLRLSRDNGLVTEY 1193

QY 1303 AYSAYCK--HKSTANDASILYNGERA-----DPVSGVTHLNGYKRS 1342
Db 1194 GYDVFORRLYKKTAKELTLFGWDGDLMIWESFKSAQTNYKHYIYEPDSFVPLLOAGYKD 1253

QY 1343 YDPTLMRHTPDSLSPFPGAGGINPYSYCLGDPI-NRS 1378
Db 1254 F---IQLIETPD-----YQEVQTKPYS-IVKDPVWNRN 1282

RESULT 9

US-09-252-991A-23640
; Sequence 23640, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23640
; LENGTH: 974
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23640

Query Match 3.1%; Score 277.5; DB 4; Length 974;
Best Local Similarity 20.4%; Pred. No. 2.5e-12;
Matches 218; Conservative 141; Mismatches 377; Indels 335; Gaps 48;

QY 504 EKFTQYRYSLIGSQSHVTLKIEERHYSATQLNSTLFOYNTDKSELGRLLKQTECTKXEN 563
Db 112 ERWYSYNALG-----LIKRADGPRTDVQDVTLYAYD-----SRGNL 149

QY 564 GKTSVVKHFTYTKQDDTLQOSHSHITTHDNFTIHRSQVRSRYTG--RLFSDTDKDIVTQ 621
Db 150 TQVTNALGQVTRLDGYDERGKPGSITDANGVTSLSA-----YTQVDCWLASVSTAGSTTR 204

QY 622 MSYDKLGRLLTRINSCTPVANTLTQYELNLOQDNRPPFVITTTDVNGNQLRNEFDGA 681
Db 205 FDYDAVGQITVTRGDG---SWLSYEY-----DDARR--LVAIGNLGERLEVDVDTK 252

QY 682 GRHVSQCLKSDGDKGFYTHITQO--YDEQGR-----HHTSTYS-DYLTNGRQOTDP 730

Db 253 GNRATQIKDASG-----LVROQWAYDELGRLLRAVAGAGGQTRSFAYDLNDNPVGETNP 308
Qy 731 DKVHLSKSYDNWGOIANTHWSYGVSEKITVDPITLTATKLOSNNSNVOTGKVTYIT 790
Db 309 RQ--FAHQAFDALDLVQSDPLGKTRLAYD-----AQNLTEVXDPKRGVTTRY 357
Qy 791 PSQOP:QITLDFRAGHL-----QSCHTLTDRGWDVRKKTDAIGQCTIYCYDNRVNI 843
Db 358 E-----YDGLGNLTSLVSPDSGTTTFEHAAGNVIRRTDARGAVTEYRYDALNRLV 408
Qy 844 QITLDPGTIVNRKYAFSTDTLITDIRVNGISLGGQTPDGLSLTQSDGGRVWAYIYSA 903
Db 409 ERSPDPSLDVGR-----YDUTADGNK-----GRLG-GLDAGAR----- 444
Qy 904 GNDQCFSTVITPDGQFIHYQYQPELDAYLQVNASNEITQOFSYNPVTGALLKAVAEGQSL 963
Db 445 -----DSLVRYPD-ERGNLVEQVRSIRLDQQLTLLDRVTRYDAA-----NQL 485
Qy 964 TPIYPSGRKLMENINDMKVSYLWTLRGLNGYDGLTGTQIKISRDTHGVTQIKDSSI 1023
Db 486 LEIGYFS-----GLAIGY-----PRNAGGQVAVS----- 509
Qy 1024 KTLTNYDDNLRHIGSQVTDLATCHMLTITVPEFDGLNRBGRKLCDSGHTLIDIQSMLKT 1083
Db 510 -TLAVGDKAPSTLVGQIAYLPFGPLRLT-----WNGITLSREYDQD 551
Qy 1084 QQLANRIV-----KLVGLOR-----TEQSYSDSRNRLNOYK-CDGASCPIDKY 1126
Db 552 YQLRQKVPWGDYQHDANGNIQHRSLWGLTDYQDPLDRUTEERGVOG----- 604
Qy 1127 GHSIVTQNTYDIYGNIT-ACHTTFADGTEDHATFKFANPTDPCOLTEVHHTHPMDPNI 1185
Db 605 -----RSVAYDAVGNRTQSDNPASGCTASSQDYQA-----PDSN 640
Qy 1186 RL-----KYDKAGRVNITDNEHNTENFTYDGLGLON-----GOGSVYGYDPL-NR 1231
Db 641 RLTAIGAQAQVTSRAG-----NLQDRA-ARKLAYDAQRLQSVLSDGQQAERYNALGQR 696
Qy 1232 LVSKQDITLDCELYRETNMLVNEVRNGEMIRLRTG-----ETIAQAQAS----- 1277
Db 697 IVKLTPESVTLYGPDGQLGEASHGSGRKLRAQYVWLDSPLATIDADYDAQKVG 756
Qy 1278 --KVLITCTDSQSVLTSDX--QNLQSOEASVAGKHSTANDAS--ILGYNGERADPVS 1331
Db 757 NPTLLVHGHDITPRLATDASQIAWQSDAFRGEALSGGQTQVNLRFPGQYDAES 816
Qy 1332 GVTHLNGVRSYDP-----TLMRFTPD-----SLSPFGAGGIMP-- 1366
Db 817 GLRY--NYFRDYDPETGRVYESDPVETLRKLNPNPENTFLNGESMLQATPYWEHGFTEPH 874
Qy 1367 -YSYCLGDPINSD-----PSGHSW-----QAMTIGMGAGLLLIAT 1405
Db 875 NYTSDNNFTAKSDKHGLSPNTDNLITYPDNCTCTLECKKTGKXSF----- 925
Qy 1406 GGVAAIAAGGIAAATASTTTALAFGLSVTSIDTSIVSGALEDASPASS 1456
Db 926 -----LVGALCKATTEF-FGVVVCNSTIWMICGASCSCQECNRAPS 965

RESULT 10

US-09-543-681A-4476
; Sequence 4476, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4476
; LENGTH: 1665
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-4476

Query Match 3.1%; Score 271.5; DB 4; Length 1665;
Best Local Similarity 20.7%; Pred. No. 1.9e-11;
Matches 233; Conservative 148; Mismatches 350; Indels 395; Gaps 66;

Qy 508 OVRYSLSQSQSHVTLKIBERHYSATQQLNSLTFQYNTKSELGLLKT-----EC 558
Db 609 QIRIDFIINQSOLVSV-KRHGIALIL-----EQDQK-LVRICSESTSPRPFVRC 659
Qy 559 TKGNGKTYSVVHKFTYTKQDPTLQOSHSTHNDFTIHSQVRSRYTGR-----LFSDDT 614
Db 660 EYDTHG-----YLSQCHAYQNH-----LWHRYSPEGLVAVMGDTD 695
Qy 615 TKDIVTQMSYDKLGELLTTLNSG-----TPYANTLTY-----DYELNNLQDNRPP 661
Db 696 STELT--IDYDEQKRVVATHSPSGFWDRIYDDYQRMYYIDABGFSRYINDN--- 750
Qy 662 FVITTT-----DVNGNQLRNEPDCAGRHVSQCLKD---SDGCKFY 699
Db 751 -LVTRTIDPLWRETVTWEQKRIAE-NEIGERTEY---GYHNGLLAVIYLPDKRAIYV 806
Qy 700 TIHTQYDEQGR--HHTSTYSYDLTNGRQOTDPPKHLSMSKSYDNWGOIANTHWSYGV 757
Db 807 -----DYNDYQOLTHFTSAFGD-----EWOLSYDENGNT----- 836
Qy 758 EKIVTDPITLTATKLOSNNSNVOTGKVT-----YTPSQPIQITLDFEAGHLOS 809
Db 837 -LVTRP---QGRQVVEYSQSGELLKAITPNGAQWQYHYNPAHLIKTT--NPYQHS 889
Qy 810 CHTLRDCGWDVRKETAIGQCTIYQ-----YDNNRVI 843
Db 890 VHS-----DELGLLHYTALNHTTRYQYSTEASTNGSLSKILLPDGVEQIDYDSERV 946
Qy 844 QITLDPGTIVNRKYAFSTDTLITDIRVNGISLGGQTFDGLSLTQSD--GGRVWAYT- 901
Db 947 AVTDCGKTRTRYCPF--DVLLAMIRPDGSEI--RPEYDSLTLRKKVNVANGSVLYER 1003
Qy 902 SAGN-----DOCESTVIT--PDGQTHYQYQPELDAVLQVASNE- 939
Db 1004 KAGQIIRVDFTGREI CVRYDRLGRRATRYPDNHELWRVY---NESGLVVEQSEW 1059
Qy 940 -----ITQOFSYNPVTGALLKAVAEGQSLTPIYPSGELKLMENINDMKVSYLWTLR 992
Db 1060 EQESRCLSTAQSYN-ARQQLIKATNPDSVVEFEYDQGRLCSELINE-QEIVHOW--- 1113
Qy 993 LENGVTDLTGHIQKI-SRDTH-----GRVTQIKDSSIKTLLNVDDLNRHIGSQVTDLAT 1045
Db 1114 ---NEADNTLALTFRGRELHYAFGALGELTSLO-----VNQAPLOFS--- 1154
Qy 1046 GEMLTITVFEGLNREIGRK-----LDSGHT-----LDIQSWLKTQQLANRIVKLVGL 1097
Db 1155 -----YNAVQCEYLRRSRAGFVNSSHYTATGLLAHORAGRGEQF-----LQSIQ 1199
Qy 1098 RTEQ-----YSYDSRNLNQYKCDGAECPDTKYGHSIVTQNTFYDIYGNITACH 1147
Db 1200 AHPQOPPECTDVHRSYQYD-----RAYNVGIE--DDRWFQT-----RYHYNANDQITETQ 1248
Qy 1148 TTFADGTEDHATFKFANPTDPCOLTEVHHTHPD--MPDNIR-----LKYKAGRVINI 1198
Db 1249 YSPQWGNODE-KFOYDNNLN---ITE-HLTPSSSMVPSDAQAMLQFQQQAGSVTR 1303
Qy 1199 TDNHNTEFTYDTLGLQ-----NGQSV-----YGYDP 1228
Db 1304 YTAGK-YQDYHYDYNGLRKLAKKIIVHTRGFRPREWRYLNTQNLCTACTPKGDCWHYTYA 1362
Qy 1229 LNRLVSQKTDLDCELYRETNMLVNEVRNGEMIRLRTGETII-----AQ- 1274

Db 1363 FGRLS-KTKVDSLAHIDPLPPQIKITWYLVSGDLIEETPIYADGTLANAQV 1421
QY 1275 ----RASKVLLTGTDSQSW-ILTSKONLSQEAAYSAG-----KH 1310
Db 1422 QWLYQGEITPTARYQCKLHVYVTHQGTPIFSEGGQASWAGRLNTWGMQFWRYRD 1481
QY 1311 KSTANDASI-----LVNGERADPVSGVTHLGNYSYDPTLMRFTPDLSLSPFG-AGGIN 1365
Db 1482 GRAENDPNTECFPFAGQVDEESGLYY--NRFYRYDRETGYLSFD---PIGLLGLN 1536
QY 1366 PYSY--CLGDPINRSPSGHLSQAWTGIGMGIAG-----LILTIAT 1405
Db 1537 PYGVVHC---PGWVDP-----FGLAGCCDELLVLAT 1566

RESULT 11

US-08-851-567B-61
Sequence 61, Application US/08851567B
Patent No. 6528484

GENERAL INFORMATION:

APPLICANT: Ernsign, Jerald C
APPLICANT: Bowen, David J
APPLICANT: Petell, James
APPLICANT: Fatig, Raymond
APPLICANT: Schoonover, Sue
APPLICANT: French-Constant, Richard
APPLICANT: Rocheleau, Thomas A.
APPLICANT: Blackburn, Michael B.
APPLICANT: Hey, Timothy D.
APPLICANT: Merlo, Donald J.
APPLICANT: Orr, Gregory L.
APPLICANT: Roberts, Jean L.
APPLICANT: Strickland, James A.
APPLICANT: Guo, Lining
APPLICANT: Ciche, Todd A.
APPLICANT: Sukhapinda, Kitixri

TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus

NUMBER OF SEQUENCES: 88

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dow Agrosciences Patent Department
STREET: 9330 Zionsville Road
City: Indianapolis
STATE: IN
COUNTRY: US
ZIP: 46268

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/851,567B

FILING DATE: 05-MAY-1997

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/063,615

FILING DATE: 18-MAY-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/395,497

FILING DATE: 28-FEB-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/007,255

FILING DATE: 06-NOV-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/608,423

FILING DATE: 28-FEB-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/705,484

FILING DATE: 28-AUG-1996

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J

REGISTRATION NUMBER: 27386

REFERENCE/DOCKET NUMBER: 960296.93804

TELECOMMUNICATION INFORMATION:

TELEPHONE: 608-251-5000

TELEFAX: 608-251-9166

INFORMATION FOR SEQ ID NO: 61:

SEQUENCE CHARACTERISTICS:

LENGTH: 1043 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-851-567B-61

Query Match

Best Local Similarity 20.4%; Score 238.5; DB 4; Length 1043;

Matches 221; Conservative 130; Mismatches 378; Indels 353; Gaps 51;

QY	612	DTDKDIVTQMSYDKLGRLLTRILNSGTPYANTLTVD---	YELNNLQDDNRPPFVIITTD	668
Db	38	DTDTR--VTRHQYDARGHL-----	NVSIDPRLYDAKQADNSVKPNFV-WQHD	81
QY	669	VNGNLRNEFDGAGRHHVSCCLKDSG-----	-----DGKFTYIHTQY	706
Db	82	LAGHALATESVDAGRTVA--LNDIEGRSVMTNATGVQRTRRYEGNTLPGRLLSVEQVF	139	
QY	707	DEGRH-----HTSYSDYLTNGR--QOTDPKVKHLSMSKSY-----	-----	741
Db	140	NOESAKVTERFIWAGNTTSEKYNLSGLCIRHYDTAGVTFLMSQSLAGAMLSQSHQLAE	199	
QY	742	---DNMGQIANTHWSYGVSEKIVTDITLTATKQLOSNNVOTGKEVITYTPSQOPIQI	798	
Db	200	GOEANWSGDETVMQGLMASEVY-----TOSTTNAIGA--LLTQTDAGNIQR	246	
QY	799	TLFEAGHLQSCHTLTRDGMWRVRKETD-----	AIGQCTIYOYDYNRNVITLTP	848
Db	247	LAVDIAGQL-----KGSWLTVKGOSEQVIVKLSWSAAGHLREHGN-----	-----	289
QY	849	DGIIVNRKVAPESTDILITDIRVNGISLGOQTFDGLSRLTQSGDGRVWAYTYSAGNDQC	908	
Db	290	GVVTEYSYEP-ETQELI-----GIT-----TTRAEGSSQSGARVL-----	-----	322
QY	909	PSTVITPDGQFIHQYOPELDDAVLOVASNEITQOFSYNPVTALLKAVAEQ-----	SLT	964
Db	323	-----QDLRYKYDPV--GNVISIHNDAAEATFRNQ-----	KVEPNRYVDSLY	365
QY	965	PIYPSGRKLMENNDMKX--SYLWTLRGLNGYTDLTGTIKISRDTGHRVTOIKDSS	1022	
Db	366	QLMSATGR-EMANIGQOSNQLPSPVIPVPTDSDTYNY--LRTVYDGRGNLQVIRHSS	421	
QY	1023	IKTLLNY-DLARNHIGSQVTDLATGHMLTTTVEFDGLNREIGRKLCDSGHTLIDQQSWL	1081	
Db	422	PATQNSYTTDIT--VESRNRVAVLSTLTPTDPRVDA-----	LFDSCGH-----QKMLI	467
QY	1082	KTQQLANRIVKLVNGVLRTEQYSYD--SRNRLNQYKCDG-----	AECPTDKYHSIVTONPT	1136
Db	468	PGQLNDWNI---RGELOQVTPVSRNSSDSEWYRSSDGMRLLYKVEQQTNGSTQVQVAT	524	
QY	1137	YDIYGNITACHTTFADGTEDHATFKFANPTPCQUTTEVHHHPDMNIRKYDKAGRVI	1196	
Db	525	Y-----LPGLELRTGVADKTIEDLOVITVGEAG---RAQVRVLHWSGKPT	568	
QY	1197	NITDNHGNTEFTYDTLGRQLONGQSVGYGDFPLNRLVSKQTDLTDLCELYYRETMVNEVR	1256	
Db	569	DIDNNQVR-----YSYD--NULGSSQJE-----	-----	589
QY	1257	NGEMIRLLRTGETTIAQORASKVLLTGTDSQOSVILTSDKQNLGOEAYSAYG-----	-----KH	1310
Db	590	-----LDSEGOILSGEYYPYGGTAIWAAR	614	
QY	1311	KSTANDASILGVNGERADPVSVTHLGNVYSYDPTLMRFTPDLSLSPFG-AGGINPYSY	1369	
Db	615	NQTEASYKFIYSGKERD-ATGLYYY--GYRYYPWVGRWLUSAD---PAGTVDGLNLRYM	668	
QY	1370	CLGDPINRSDPSG-----HLSWQAWTIGMGIAGLLLTIATG---GMAIA-----	AAG	1414

Db 569 VRNPITLTDHGLAPSPNPNRNTFWASFLPRKPDCEGMSASMRGQKIGRAIAGGIAIG 728
 QY 1415 GIRAIASTSTAL--AFGALSVTSDITSIVSGALEDAPKASSILGWVWGM-----1465
 Db 729 GLAATIAATAGAAIPVILGVAAVAGAGIGALMGYNVGSLLERGAALLARLVQKSTLVOSA 788
 QY 1466 --GAAGLAESAIGK--GFKATHLGAFAEDGENALLKSTSESSRIKWGVTRSLDREIV 1519
 Db 789 AGAAGASSAAAYAGAAQGVVASAAGAVT-----GAVGSWINNAD 829
 QY 1520 RNEGGVQIKHDSRYTNDNFWGKEQAILVH-----GDKGFLYHTGNK-----HNCKG 1568
 Db 830 RGIGGAIGAGSAGVTITDMLGTA--STLTHEVGAAGGAAGGMITGTQGSTRAIGHAGIG 887
 QY 1569 PY 1570
 Db 888 TY 889

RESULT 12
 US-09-328-352-5545
 ; Sequence 5545, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; PRIOR FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 5545
 ; LENGTH: 804
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 ; US-09-328-352-5545

Query Match 2.7%; Score 235.5; DB 4; Length 804;
 Best Local Similarity 20.8%; Pred. No. 3.3e-09;
 Matches 147; Conservative 92; Mismatches 253; Indels 215; Gaps 30;

QY 699 YTHITQYDEQBHHHTSYSDILNCRQOTDQKHLMSKSYDNWGOIANTHS-----753
 Db 21 FTLQTEYDAQGAIRSENTQAGRVTKYKYE-----VSFSTCL---GNLASFQWSKVFSC 71
 QY 754 ---YGVSEKITVDPITLTATKQLQSNNNVQTKVTTYSQ--PIQITLDFEAGHLQ 808
 Db 72 LSNVVLGDQHVGLPIEV-----DVNGKEITQFAGYKRGIPQVKLANGA---116
 QY 809 SCHLTRDGDWRKRDGAIQCTIYOYDNYNEVIQITLPQGTIVNRKXAPFSTOTLTD 868
 Db 117 -TETNIVDDPGNITQHTDAGVSRKQYDDAGRL---YIDPPIVGLNYSFTFYDGTJVS 171
 QY 869 IRVNG-----IS---LQOQTFD-----GLSRLTQS- 890
 Db 172 RVVTGGQLSRIEKNGDGLLSSBCKISNKSINSKYDAFGNLIKSNPGFAITSGT 231
 QY 891 -----QDGRWYATYSAGNDQCPSTVITPD-----GQPIH---921
 Db 232 TSSYVDFDRPITVNDNGSVVTCYOSCGGTGAIVQTTDSFGTTESNLLAAGDFSALKT 291
 QY 922 -----YQYQPELDDAVLQ---VASNEITQQFSY-----NPVTGALLKA 956
 Db 292 LVARKGTGDSVFQTTTEFENALLKPKVAVSGSSTQSYTNGNTTILATEKNSISGQKTFK 351
 QY 957 VASQSLTPIYPSGRKMNENDMKMSYLWTLGLENGYT-DLTGFIQKISR-----1009
 Db 352 YDGTGRTIITHPDSSVETIKVFQKDLASHTWREVTTSYSLAGRLKTTTIANISEA 411
 QY 1010 ---DTHGRV-----TQIKDSSIKTTLNDDYDLNRHIGSQVTDL-----A 1044
 Db 412 FDLDTYGRVISHQKINANDTNNVYVRYGNQLN-----QVTSIQYENGKSVNLNQNA 466

QY 1045 TGHMLTTTVEFDGLNREIGRKLCDSSGHTLDIQQSWLKTQCLANRIVKLVGLQR-----TE 1101
 Db 467 LGEVTSIPNVIOSLNARNARQQLTTVQANTDTLWSYTTNDSSLNLSATS--ERKCVLVN 524
 QY 1102 QYSYSDRRNLNOYKCDGABCTDKYGHSVTONTFYDIYGNITACHTTFADGTE-DHATF 1160
 Db 525 DYGYDKLNRVNLK-----SDKCGS---VYNATIDRYG-----TCLMSTVELDQARY 567
 QY 1161 KEANPTDPCQTEVHTHTPMDNIRLK-----YDKAGRVINITDNGHNTENFTY 1210
 Db 568 QYGYND--DIKV-----NITSSTVAPAITVYNTVNTSRLASVSGSTISFTY 616
 QY 1211 DTLGRLONGQSGVGYDPLNRLVSKQTDTLDCELYRETMLVNEVRN 1257
 Db 617 DAMGNVNDGVRATLTLYDYSRLNKNENY---LYNADGLRVRVARD 660

RESULT 13
 US-09-252-991A-31032
 ; Sequence 31032, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 31032
 ; LENGTH: 1128
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-09-252-991A-31032

Query Match 2.5%; Score 223; DB 4; Length 1128;
 Best Local Similarity 19.6%; Pred. No. 5.6e-08;
 Matches 258; Conservative 151; Mismatches 467; Indels 442; Gaps 68;

QY 235 NTSIGNENPLTWSFGYTP-----IGNGLIGOWITSMTPAGLAKETVYNNNOGHH--FP 288
 Db 63 DPALAAPDTFTFARGYLLSNPRIGRLG-RGWLPFGSMHLELGEDACVLVDQORRIGFP 121
 QY 289 QSANLPVLPVYVTLMKQVPGAGQPAIQAEYSYTSHTYVGGSGNGIWNKLNLYGLATEYN 348
 Db 122 ALA-----PGA-----QHYS-----GSEELM-----137
 QY 349 YGSTERRYKKEGHQDQIVRIERTYNNVHLLTSECKQONGYIQTETAYVAITGHNFDSQ 408
 Db 138 -----LRKGGSGGGEQAQR-----GRWAAVPAB-----LOTQEGSVLVLSGHSY--177
 QY 409 PSQFOLPKTKTETWRSADNSYRSBIT-----ETTFDESGNPLTKVTKDKK-----TQ 455
 Db 178 -LHFQ-----RCPDGIWRLOASFGRAGYRTFEFWSGRGLLTGVRDSAGRSYALVYQ 227
 QY 456 KLIISSTHWEYPPAGEVDCNCPPEYGFTRFKXIIQTPYDSFKD--DPEK-----FIQ 508
 Db 228 QACEPS-----EGDD-----GLRUF--GVILLASHDGPFPYIDPQSGPLDMLVR 269
 QY 509 VYSLIGSQSHVTLKIBERHYSATQLL-----NSTLFOYNTD-KSELGRLLK 554
 Db 270 YQFSDSGD---LIAVRDLQVVRVFAWRHMLVAHCEPGGLEVRVYEWVHAPHGRVVK 325
 QY 555 QTECTKGNGKTYSVVHKFTYTKQDDTLQOSHSTTHDNTIHRSQVRSYTCRLFSDDT 614
 Db 326 QIEAG-----GLTRTFYLR--DATEVSDSLGRVERFEFAGEGGRWTLVLRADGS 375
 QY 615 TKDVIQMSYDKLGRLLTRTLNSGTYPANTLTVDYELNNIQQDNRPPFVITTTDVGNNQL 674

Db 376 R-----SEFDYDLFGRL-----VAMRDPLGRET 398
QY 675 RNEPAGRHVSQCLKSDGDKFYTHIQYDEQGRHHTSYSDYLTNGRQOTDPDK-- 732
Db 399 RRRRDGQGRMLEE-----ESPGKARY-----RKRVDDEETGL 429
QY 733 -VHLSMSKSYDNWQGANTHWSYGVSEKITVDPITLTATKOLQSNNVQ--TGKEVTY 789
Db 430 LVELE-----DAMQRRWTFERDER-----GNATTVRGFAGSTRAY 465
QY 790 TFSQOPIQIT-LFDAGHLOSCHTLTRGWR--VRKETDAIGCCTIYQVNDNRVIOI 845
Db 466 EDRPLPDAPTRIVDPGRGER-----RLEWRNFGLLAALATDCSGQWRYDYDNEGRIVAS 519
QY 846 TLPDGTIVNRKYAPSTDTLITDIRVNGISLGQOTFDGLSRLTOSQDG-----G 894
Db 520 SDPLGQLFRYRVDPLG--QLIGLELADGALSABE-YDALGQOTRIADAEHATLFSWGHG 576
QY 895 RWANTYSAG-----NDQCPSTVITPDGQFIHYQYQPE-LDDAVLQVASNEITQOFSY 946
Db 577 DLLARVSDAGGELSILHDEAGRLVALTNENGVOAQFRYDLRLVEETGDFGRQRYR 636
QY 947 NPVTGALLKAVAGOSLPIYVPSGRKKNENINDMKM-----SYLWTLRG----- 992
Db 637 NAADELIAREDAGRETTYADROGRLASIRVPATEHAPALVERVWMLADGLASAGAD 696
QY 993 LENGVT-DLTGIQKISRDT--GRVTOIKDS-----SIKTTLYNDDLRHIGSOVTDL-- 1043
Db 697 CEVRYTYDEVGNL-RLESOVHADGWYSVEHSHDALGVQRTSYGD-----APPVAMLY 750
QY 1044 ATGHM--LTTTVER-----DGLNREIGRKLDCSSGHTLIDIOQSWLTKQ-- 1088
Db 751 GPHLHGALVGAVALAFERDALHREVRDA-----RRDQDDALFTQERQHAPLGLR 804
QY 1089 RIVKLVQIQRTEOYSDSRNL-----NOYKDCGAECPDKYGHSI-----V 1131
Db 805 SRLLAGGDFWQGVRYDGLQVAGIDNQY-----PSVRYEYDLCGLLASRRAGAA 857
QY 1132 TONTYDIYGNITACHTTFADGTEDHATFKFANFTDPCQLTEVHTHTHPDNDIRLYDK 1191
Db 858 ASTRYDAAGN-----RLEGVGEYAR-----EDARQAFANELYRSFGRSEVRANQ 904
QY 1192 AG-----RVNITDNHNGTENFTYDTLGR--ONQGSVYGVYDPLNRLYSQKT 1237
Db 905 AGEQFARWAGNRVERIAGN-----RYRFDALGNLVERIGADGERLRLAYDGAORLVHLTR 959
QY 1238 DTLD--CELRYRETMLNVEV-----RNG--EMIRLLRTGETIIAQORASKVLLTGTDSQ 1287
Db 960 DYADGTREARYRYDALSRRIAKVLRDQVEQVRFQWGDGDRQCAEAFARELATTTVEPG 1019
QY 1288 QSVIL-----TSDKQNLQ--EAYSAYGKHKSTANDASILYNGERADVPVSGVTHLGN 1338
Db 1020 GFVPLLRLQACEPDPPELQLRQAFAGQPL-----PAQCVPALGE 1062
QY 1339 GYRSYDPTLMRFHTPDSLSFPFGAGGINPYSVCLGDPINRSDPSGHLISWQA-----WTGI 1392
Db 1063 ARIAF-----FHTDH-----LGTPLQLSDERGQLRWQGVDPDWRV 1098

RESULT 14

US-09-328-352-5486
; Sequence 5486, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5486
; LENGTH: 3290

TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-5486

Query Match 2.5%; Score 222.5; DB 4; Length 3290;
Best Local Similarity 18.8%; Pred. No. 3.9e-07;
Matches 370; Conservative 249; Mismatches 656; Indels 697; Gaps 101;

QY 12 ITMSDNEF-----FTQANNFTSAVG-----GVDPR 39
Db 459 LTTSLDNDTGVSATDQITQKXFNLEKGEQTSQVTVLSTDEGKTWQETTVAQKDLAD 518
QY 40 GLYNIQITLCHIVGNGN-----LCPTLPLT--LSYSPLNKTDIGFGIFNGFLGSVD 89
Db 519 GIYQYKAVTDVAGNISETAIOKVVDVTTAPQACKLTLSLNDT-----GVSATD 568
QY 90 ---RKNS-----LLSSTGE-----NYKVIETDKTVKLOKKLDNLRFXDKLKENCYRII 136
Db 569 QITQDNSFTLKLQPIVIGEQAALLDHYEVSDEGKTWQETTAD-----QKLDADGIYQ-- 622
QY 137 HKSGDI:EVLTGFNNAFDLKPKKLNPAAGHAIYIDWNEFATQPLNRIVDLDGHDIFL 196
Db 623 YKA-----VITDLAGNISESAIOKVVDNS-----LNVESTTVIVKPTED----- 663
QY 197 LNLVYQCLIKTILTFPGQKEGYRTELRFNRLQNSIHNFLGNENP-----LWTSFGYTP 252
Db 664 ---NTISLVEKDQVISIRLBIANLPTDLNS-----SLTSVNTLENVTYNPHFDE 710
QY 253 ICKNGILGOWIT-----SMIAPGGLKETVYNNQGHHPFQSAN 292
Db 711 VTQ-----EMWTEIPAEFLWSVEPQTNISIDISLTDAQNTAITHTQN---YNDPTFN 762
QY 293 LPVLVTVTLMKQVPGAGPAIOAEYSYTSHTNVGGSGNGIWNKLDNLYGLMTEYNYGST 352
Db 763 SPILDSLT--FNLDGA-----LISGSAYKSKVDIYNKNGDWLASTIT----- 804
QY 353 EERRYKDEGH-----DOIVRIERTYNNYHLLTSECKOQNGYI--QTTETAYAI-- 400
Db 805 ---NEEGKFTLODLSINTNQEVAVATNGY-----SSNSSIGLAVTEVPAISIT 851
QY 401 ---IGHFDSQPSQFOLPKTKTETWESADNSYSEITETTFDESG----- 442
Db 852 RISPEGVISGYATESHFIV-----KQNGNLQEFNSNVFDSGITPFSVWALGEV 903
QY 443 ---NPLTKVKDKTKTKIISP-----STHWEY-----YPPA-----GEVD 474
Db 904 RPFILSLDQPL-----EEGAQIIISTDKDNISCHQYITADYTPAVFLETQFDISGETL 958
QY 475 NC-PEPYGFTFRVK--KIIOTPYDSFEEKDPEKFIQYR-YSLIGSQSHVTLKIEERHY 529
Db 959 SVHVNEPNSFIRAFSGEGLNIATGF-----TDEQGFASLQVFPQFKSGETVTVQVVDKNQ 1013
QY 530 SATQLL-----NSTLFOYNTDKSELGRLLKQTECTKGE 562
Db 1014 NTSSETLIEVPNFAYIPHVERITQEGLSGVAEDNSTVIVROADGNELKV----- 1063
QY 563 NGKTSVVKHFTYTKQDDTLQOSSHITHTDNFTIHRSOVRSYTG-----R 608
Db 1064 ---TLGDD-----NSWSDFGHFSLSVNRPLIDGEEKISVQIIDNKGLMSPEQN 1107
QY 609 LFSDDTDKDIVTQMSYDKLGRLLTTLNSGTPYANTLYD-----YELNNLQDD-- 657
Db 1108 IIVDLTPPPAPTELNFNDAGLV---YGHAEFFSEILVKDQGNILNKFNNWNTDESGS 1164
QY 658 ---NRPPVITTTDNGN-----QLRN-----EPDG-----AGRHV 685
Db 1165 FSIELGTFLTNAETVYVVTATDVNGNVSLAAQIAPNAPAFYVDSFTSDGVISGOAENS 1224
QY 686 SOCLKSDGDKGKFTYTHIQYDEQGRHHTSYSDYLTNGRQOTDPDKVHLSKSYDNW 745
Db 1225 TLVWAKAGD-----VVAEIKVGEONGWNGSSIFK--LQDLRPLVDGEQFFLSIK---DARG 1276
QY 746 QIANTHWSYGVSEKITVDPITLT--ATKQLQSNNVQVG--KEVTTYTPSQOPIQITLFE 803

Db 1277 QVS-----ADVTITADTAPASNLVSEDSGLTVAELNT-----TIQVFDH 1321
Qy 804 AGHLQSC--HTLTRGDNDVRKETAIGQCTIY-----QYDYNRVIQ 844
Db 1322 NGQLVNIWNNTNSD-----GTFIYLGNNLHGEAFTVTKDQAGNVSEALS 1369
Qy 845 ITPLDGTVIARKAPFSTDLITDIRVNGISLGO-----TFGLSELTSQSQGRV 896
Db 1370 INAPLDDI-----APNPKILLD--ANGQNTAQAEANSQIEVFDSLQ-----NQIG--- 1415
Qy 897 WAYTYSAGNOCPSVTITPQGFTH-YOYQPELDDAVLVASNE-----ITQOPSYN 947
Db 1416 WGSTDSAGN-----ASGSFNQTYLHGEELTFVVIDRAGNRSIBFKQNALIDTIAPN 1466
Qy 948 PV-----TCALLKAVAGOSLTPIYPGSLKWNENDMKMSYLTWLRGLENGYTDL 1000
Db 1467 PIENIIFNENGQSTAQAEAGSSIDVLDQIG-----NKIGF-----GYTDS 1507
Qy 1001 TGTQIKISRDIT--HG-----RVTOIKDSSIKTLLAYDLDNRHIGSQVTDLAT 1045
Db 1508 SGNVSGYFQOYVYLHGEELTFVVIDRAGNRAEVKQSA-----LNDVAPNPIENILLDL-N 1562
Qy 1046 CHMLT-----TTFEFDGLNRE-IGRKLCDSSGHTLD-IQOSWLKTOOLANRIVKLGVL 1097
Db 1563 GQNTAQAEANSQIEIKNNNGDVVGYGSADSNVSGYLYOVHLHGEELTFVVDLAG-- 1620
Qy 1098 QRTQSYSDSRNRLNOYK-----CDGAECPTDKYGHISVITQNET----- 1136
Db 1621 -----NRSTEVKQNALIDDIAPNPIENIVLDINGQNTAQAEANTQIEVKNVAG 1669
Qy 1137 -----VDIYGNITA-CHTTAOGTGDHATFKKAN-----PTDPCQLTEVHHTHDMPD 1183
Db 1670 EIVGSGYVDGAGNVSGYLYQVYLHGEELTFVVIDRAGNRSIEVKQNALIDDIAPNPIE 1727
Qy 1184 NIKLYKDRAGRIVINTDNGHNTENFT-----YDTLGRQLONGQSGYVGYDPLNRLVSKQTD 1239
Db 1728 NILL-----DANGQNTAQAEANTQIEVKNVAGVIG-----SGSTDS 1765
Qy 1240 L-DCELYRETMLVNEVANGEMIRLL-----RTGETIIAQQRASKVLLTGTDSQSQSVILTSD 1295
Db 1766 MGNVSGYFQVQVYL-----HGEELTFVVIDRAGNRSIEVKQNALIDDIAPNAIENIIFEN 1820
Qy 1296 KQNLQSEA-YSAKGHKSTANDASITLVNGERADPVSG-----VTHLGN 1338
Db 1821 GQNTAQAEANSKVEVKNVAGVGSY-VDSAGNVSGYLYNOVYLKGEELTFVVIDQAGN 1879
Qy 1339 GYRSYDPTLMRFHTPDSLSFPGAGI-----NPYSYCLGDP-----INRSDPSGHL----- 1384
Db 1880 --RSIEVQKTA--LDNTAPENATNLVSEDSGLSGMAEPNATQIFDOVQQLLNOWN 1935
Qy 1385 --SQAQWIGMG--IAGLLLTATGMAATAAGGIAAATASTTALAFGALSVTSDI 1439
Db 1936 NYNWDGTTNLYNSNYMEGEVFKV-----VVVDQAGNLSGEV----- 1973
Qy 1440 TSIVSGALEDASPKASSILGWVSMGAGLAESAIGKGTKLATHLGAFAEDGENALLKS 1499
Db 1974 ---VKAPLDDTAPVAASDLVFNEDSSLSGVAEP-----NTFIQIFDQNGQQ--MNT 2020
Qy 1500 TSESSRIKWVTRSLDREIVNEBQVVKDHSRGVTDNFMKGKGCQAILVHGDKGFLYHT 1559
Db 2021 WSOQSVNADGTFITFFGTYNLHGEETFTV-----VKDLAGNVSEAVSV----- 2062
Qy 1560 EGNKENGKGPYTRHTPEOLVDYLDKNNIVDITQGGD--KPVHLLSCYKSSG 1609
Db 2063 -----KAPLDDIAPNFIKNIIVFDANGQSTAQAEANSQIEIFDSFGSQIG 2107

RESULT 15

US-08-447-031A-2

; Sequence 2, Application US/08447031A

; Patent No. 5851794

; GENERAL INFORMATION:

APPLICANT: GUSS, Bengt
APPLICANT: HOOK, Magnus
APPLICANT: JONSSON, Hans
APPLICANT: LINDBERG, Martin
APPLICANT: PATTI, Joseph
APPLICANT: SIGMAS, Christer
APPLICANT: SWITALSKI, Lech
TITLE OF INVENTION: A COLLAGEN BINDING PROTEIN AS WELL AS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/447,031A
FILING DATE: 22-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/861,804
FILING DATE: 21-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/SE91/00707
FILING DATE: 22-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: SE 9003374-7
FILING DATE: 22-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 012889-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1183 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-447-031A-2

Query Match 2.5%; Score 220.5; DB 2; Length 1183;

Best Local Similarity 19.1%; Pred. No. 9.6e-08;

Matches 206; Conservative 159; Mismatches 405; Indels 311; Gaps 56;

Qy 299 VILMKQVPAGQ---PAIQAEYSYTSNHYVGGSGNINNNKLDNLYGLMTEYNYGSTESE 355

Db 205 ITIKDIOGGQQLDLSTLNINVTGTHSNYSQS-----AITDFEKAPGSK 251

Qy 356 RYKDEGHQDIIVRIERTYNNYHLTSECK-----CQNGYIOTTETAYAIIG----- 402

Db 252 ITVDNKTNTIDVTIPQGYGSYNSFSINYKTKITNEQKEFVNNSQ-AMYQEHKEEVNKG 310

Qy 403 -----HNFDSQPSQFQLPKTKTETWRSADNSYRSEITETTFDEGNPLTKVVKDKTKQ 456

Db 311 SFNHTVHNINANAGIEGTVGELKVLKQ-DKDTKAPIANVKELSKD-GSVVKD--NOK 366

Qy 457 IISPSSTHWEYPPAGEVDNCPPEPYGFTFVKIITQTPVDSFEKDDPEKFIQYRSL--I 514

Db 367 EI-----EIIIDANGIANIKALPSG--DYILKEIEAPRPYTFDKKE-----YPTMKDT 414

Qy 515 GSQSHVTL-----KIE-----ERHYSATOLLNSTLF-----QYNT---DKSELGR 551

Db 415 DNQGYFTTIENAKAIEKTKDVSQAQVWEGTKQVKTIFYKLYKQDDNQNTTPVDKAEIKK 474

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2004, 05:44:38 ; Search time 67 Seconds
(without alignments)
7772.803 Million cell updates/sec

Title: US-09-889-874A-23

Perfect score: 8879
Sequence: 1 VIKFLKLFRRITMSDNNF.....PRKIILGRTEKTVKPKTRP 1673

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 1276540

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCT05_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	932.5	10.5	982	12	US-10-365-742-106
2	517.5	5.8	1317	15	US-10-369-493-11243
3	420.5	4.7	2364	14	US-10-156-761-7834
4	395	4.4	820	14	US-10-156-761-7990
5	392	4.4	1250	14	US-10-156-761-7572
6	391.5	4.4	2386	14	US-10-156-761-7751
7	386	4.3	2234	14	US-10-282-122A-46565
8	381	4.1	1510	15	US-10-418-861B-55
9	353.5	4.0	843	12	US-10-282-122A-76214
10	350	3.9	1515	12	US-10-282-122A-47600
11	342	3.9	1397	12	US-10-282-122A-43059
12	340.5	3.8	1411	12	US-10-282-122A-43060
13	339	3.8	1426	9	US-09-912-020-340
14	339	3.8	1426	12	US-10-282-122A-42617
15	331.5	3.7	1377	9	US-09-815-242-10384

16	331.5	3.7	1377	12	US-10-282-122A-42731	Sequence 42731, A
17	331.5	3.7	1377	14	US-10-287-274-467	Sequence 467, App
18	328	3.7	1572	12	US-10-282-122A-69415	Sequence 69415, A
19	322	3.6	1385	12	US-10-282-122A-68242	Sequence 68242, A
20	317.5	3.6	1565	12	US-10-282-122A-44558	Sequence 44558, A
21	314	3.5	1395	12	US-10-282-122A-50586	Sequence 50586, A
22	307.5	3.5	932	12	US-10-282-122A-44617	Sequence 44617, A
23	306.5	3.5	1530	12	US-10-282-122A-69175	Sequence 69175, A
24	298.5	3.4	1531	12	US-10-282-122A-49308	Sequence 49308, A
25	297	3.3	1512	12	US-10-282-122A-78386	Sequence 78386, A
26	291.5	3.3	1253	12	US-10-282-122A-68914	Sequence 68914, A
27	291	3.3	2346	12	US-10-072-012-491	Sequence 491, App
28	290.5	3.3	2802	9	US-09-808-602-81	Sequence 81, App
29	290.5	3.3	2802	10	US-09-800-198-69	Sequence 69, App
30	290.5	3.3	2802	12	US-10-072-012-489	Sequence 489, App
31	289.5	3.3	1364	12	US-10-282-122A-75309	Sequence 75309, A
32	285	3.2	2613	16	US-10-038-854-42	Sequence 42, Appl
33	285	3.2	2628	16	US-10-038-854-40	Sequence 40, Appl
34	285	3.2	2721	16	US-10-038-854-38	Sequence 38, Appl
35	285	3.2	2725	16	US-10-038-854-36	Sequence 36, Appl
36	284.5	3.2	2715	12	US-10-042-865-52	Sequence 52, Appl
37	284.5	3.2	2715	12	US-10-029-020-51	Sequence 51, Appl
38	280	3.2	1317	9	US-09-815-242-5118	Sequence 5118, Ap
39	280	3.2	1317	12	US-10-282-122A-43495	Sequence 43495, A
40	279.5	3.1	944	12	US-10-282-122A-47806	Sequence 47806, A
41	278	3.1	985	12	US-10-282-122A-43061	Sequence 43061, A
42	276.5	3.1	1438	12	US-10-282-122A-78360	Sequence 78360, A
43	273.5	3.1	1688	15	US-10-144-194A-113	Sequence 113, App
44	273.5	3.1	1737	9	US-09-808-602-83	Sequence 83, Appl
45	273.5	3.1	1737	10	US-09-800-198-71	Sequence 71, Appl

ALIGNMENTS

RESULT 1

US-10-365-742-106
; Sequence 106, Application US/10365742
; Publication No. US20030204868A1
; GENERAL INFORMATION:
; APPLICANT: Collmer, Alan
; APPLICANT: Alfano, James R.
; APPLICANT: Cartinhour, Samuel W.
; APPLICANT: Schneider, David J.
; APPLICANT: Tang, Xiaoyan
; TITLE OF INVENTION: PSEUDOMONAS AVR AND HOP PROTEINS, THEIR ENCODING
; TITLE OF INVENTION: NUCLEIC ACIDS, AND USE THEREOF
; FILE REFERENCE: 19603/4112
; CURRENT APPLICATION NUMBER: US/10/365,742
; CURRENT FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: 60/356,408
; PRIOR FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: 60/380,185
; PRIOR FILING DATE: 2002-05-10
; NUMBER OF SEQ ID NOS: 209
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 106
; LENGTH: 982
; TYPE: PRT
; ORGANISM: Pseudomonas syringae pv. tomato DC3000
US-10-365-742-106

Query Match 10.5%; Score 932.5; DB 12; Length 982;
Best Local Similarity 29.7%; Pred. No. 1.1e-65;
Matches 285; Conservative 142; Mismatches 357; Indels 175; Gaps 33;
QY 642 ANTLIDYELNNLQDDNPPFVITTTDNGNOLNEFDGAGRHYVSOCLKSDGDKFYTI 701
DB 31 SSTLGARYD-----DWNOR--CCVTEYSDPIGSDVHK-----GPTQKT 74
QY 702 HFOQVDEQGRHHTSYSDYLTNGRCQTDPPDKVHLSMKSVDNWCQIANTHWSYGVSEKIT 761
DB 75 WKQSGDPEGR-----ISGRSET-----WLNL-----FKPDRIR 103

QY 762 VDPITLITATKQLOSNNVOTGKEVYTYTPSOQPIQITLFDAGHLQSCHTLTBDGWDV 821
Db 104 -----TITAGTGRSHRSRNLTT-----TEQEL-----SRQFLYDGLGRC 144
QY 822 RKETAIGQCTIYQYDYNRVITQITLPDGTIVNRKYAPFSDTLITDIRV---NGIS--- 875
Db 145 TEQDALQOSTLFSYDNWSRVSSTLADGSLVNRSYAFQSSSELATMLEVVHONGTTRTV 204
QY 876 LGQCTEDGLSRLOSODGGRWAYYSAGNDQCPSTVITPDQFIHYOYQPELDDAVLOV 935
Db 205 AGTKQFDEGSRVOTGTGRVSOQFNVDGEMO-PRSRITAGLDNINFTYTRALTDQIFSS 263
QY 936 ASNEITQOFSYNPVTGALLKVAEGOSLTPPIYPSGRILKMNINDMKMSYLM-----T 989
Db 264 TAPDETAKFDYKTSARLIEATNPQGTTRYVDVHNLQTLGETWDL--LQOAMETHOSS 321
QY 990 LRLENGYTL-----TGFIQKISRDTHGRVTOIKDSSIKTTLNVDLNRHIGSOVTDLA 1044
Db 322 LLGRPIKRDLLKGEAGAEIRYDITLGRIFNQSRLRTIDYDVLGQCKVATEDLQ 381
QY 1045 TGHLMTTTFEFDGLNREIGKLCDSGHTLDIQOSMLKTOOLANRIVKLVGLQRTQYS 1104
Db 382 AGTVIIDMEYDDQGEILRTQASNOAALTQTWAVDGLLKTDRDQOAGSPILLHETFS 441
QY 1105 YDSRNLNQKCDGAECPDKYGHISVTCNFTYDIYGNITACHTTEADGTEDHATKFEAN 1164
Db 442 YDPRGLTLVNYLGSPLRDELQRETRQIFSDBLDNITICOTRFTDGTISERAAFIKS 501
QY 1165 P-----TDPCLTEVHTHPMDNIRLKYDKAGRVINITDNGNTEFTYDITLGR--- 1216
Db 502 PGDKKHKRCOLLSTAYTPRKTPTPTSDYANGN--QLKDEHGN--SLHYDSOSRLLOV 557
QY 1217 -ONGQ--SVYGDPLNRLVSKTTLDCEL--VYRETNVNEVRNGEMIRLLRTGTII 1271
Db 558 AETGGAPIQYVDGHNQVATR-DONESEILRFEGHQLSSTVOEDORTYHLGEOPL 616
QY 1272 AQO---RASKVLITGDSQSVILTSKQNLQSEAYSAYGKHSTANDASILVNGERAD 1328
Db 617 GQOIVDDAEQTLILLTDANQSVMGFEQOQLKAVYSAYGERHSEALLSTAGENGVE 676
QY 1329 PYSVTHLNGVSEYDPTLMRFEHPSLSPFCAGINPYSYCLGDPINRSPSGH----- 1383
Db 677 AANGWLLNGVRAYNPLMRHSPSLPFPRAEGVNPYICLGNPIALRDPDTHDASQ 736
QY 1384 -----LSWQ-----ANTGIGVIA-----GULLTIATGGMIAAAGGI----- 1416
Db 737 TGLRPRFDEGALPMQGGGDMGWGVGTGVVFTVLGVAATITLGTATPTVGPVTLGI 796
QY 1417 -----AALASTSTALAFG-----ALSVTSDITSIVSGALEDASPKASSILGHVS----- 1462
Db 797 SMTASAAAASVSTGALLIVGTALTAATTANTVAIVNN-----DQTAGVGHGIAAV 851
QY 1463 -----MOMGAAGLAESAIGKGTILA-----THLGPASDGENALLKSTSESR 1505
Db 852 PVGLVFGAGAVVARAAAKVAAANAGTIGVRSVRIG-LAAAGARRTITSAASAR 909

RESULT 2

US-10-369-493-11243
; Sequence 11243, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11243
; LENGTH: 1317
; TYPE: PRT
; ORGANISM: Methanosarcina mazei
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(1317)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-11243

Query Match 5.8%; Score 517.5; DB 15; Length 1317;
Best Local Similarity 21.2%; Pred. No. 8.6e-32;
Matches 334; Conservative 194; Mismatches 505; Indels 541; Gaps 72;
QY 39 TGLYNIQITLGHVGNGLGPTLPLTL--SYSPLNKTDIGFGIG--FNFGLSVYDEKNSL 94
Db 56 TGSFYQYQDLSIFRG-----LPLTVSRYSNWDNRSGLFSGSWTFNVMNLTVDNNGN 110
QY 95 LSLSTGENYK---VIETDKTVLQOKKLDNLRPEKDKENCYRIHKSGDIEVLTFGNNN 151
Db 111 VTVLGGDGHDTYILNPDGTVSRPLSVFDDL-----IKNSDETTLTKKDOT 157
QY 152 AFDLKVPPKLAN--PAGHAIYIDWNEATOPRLNRIYDLDLGHDIPLNLEVOGLIKTI 208
Db 158 KYNFSSEKLVNIYDKNQNI---NFTYTCEQLTKV--TDASREL--ILAYDHG----- 206
QY 209 LTLFPQKEGYVTELRLFNRLQNLNHNFLSGNENPLTWSFGYTPIG-----KNGILGQ-- 261
Db 207 -----HIISITDPMGRVMSYSYDDQGNLQCKNPIGGKLS 241
QY 262 -----WITSMTAPGG---LKETVNSNNQGHHPQSANLPVLPVTLMKQVPGAGQP 311
Db 242 VTYDENHMTSITDPGRNHPKNTY-----DEKGVLSQS----- 276
QY 312 AIQAEYSYTSNHYGGGSGNGLNKNLNLGLMTEVNYGSTRRYKDEKGDQIVRIER 371
Db 277 -----NSLNATY-----TFNYDS--ENRK----- 293
QY 372 TVNNVHLTSECKQONGYIOTTETAYVAILGHNFDSQPSQFQLPKTKTETWRSADNSYS 431
Db 294 -----TTETD-----PFCNKYTSFDEHFW-- 313
QY 432 EITET-----TFDSGNPLTKVTKKTKOKIISPSHTHWYPPAGEVNCPEPYG 482
Db 314 ELNETNQLGYTISAYDENGNRISVTNENSKTKLA-----YDANGNIK--TTNPLG 364
QY 483 FTRFVKIIQTPYSEFPKDPKPEFIQVRYSLIGSQSHVTLKIEERHYSATQLINS----- 537
Db 365 YSK-----SMTYDS-----KNNLISQTDLRHKTSEFYDDNSNLKSIDALG 406
QY 538 --TLPOINTKSELGRLLKQTECK-----GENCK-----TYSVVHKT 574
Db 407 HETVFSYD-----KYGVIGETDSNKKTATFSYNNNGDQITIDANGKTSFTYDTVGRVT 462
QY 575 -----YTKQDDTLQOQSHSITTHDNFTIHRSQVRSRYTGRL-----FSDTDTKDI 618
Db 463 TKTDAGNRYTFDYDALDNLISITD-----PVGQTTSTNTMLETKLVQLMKVXDQSI 516
QY 619 VTQMSYDKLGRLLTRTINSGT---PYA-----NLTLYDYELNNLODNRPFVIT 665
Db 517 LTLIXSLKXKXQMWEXLGNMMPLAINPPTOKGHKTSYDYDELNQ-----VS 568
QY 666 TTVNGNQLRNEFDGAGRHSV-----QCLXDSGDGKFTYTHIQOQYDE 708
Db 569 VTNALGKTRNKYDAIGNKISITNAYGKSTRYSYNSLNLQVKTVMAMGK---VVRYNYDA 625
QY 709 QGRHHTSYSD-----YLTNGRQOTDPPDKVHLMSKSYDNWNGQIANTHWSYVSEKIT 761
Db 626 VGNLISITDENGHKINGYDSLRNQSVTDALRKTTRNKYDAVGNKISITNAYGKSTRYS 685

QY 762 VDPITLTKATKQSQNNVOTGKEVTTTPSQOPIQITLFDAGHLQSCHTLTRDGMWRV 821
Db 686 Y-----NSLN-----QLVKVTAMGGVVRVYNDVAVGNLIS-----715
QY 822 KETDAIGCTIYQDYNRVIQITLPGDTIIVNRKYAPFSTDTLITDIRVANGISIGOOTF 881
Db 716 --TTDANGRKINYGDSLNRQVSIINALGKTRNKYDAVGNKISSTDANWR--LTKVSY 770
QY 882 DGLSRLTOSQDQ-GRVWAYTYS-A-NDQCPSTVITPD--GQFIHQY-----QPELDDAV 932
Db 771 DSLNRLVKVTAMGGVVRVYTDVAVGN-----LISTDAKGEKTDYGVDSLDRQVSIITDPL 825
QY 933 LQVASEITQOFSYNPVTGALLKAVAEGOSLTPYIPSGRLKMNINDMKMSYLWLIRG 992
Db 826 GRTARK-----YDAVGNKISSTDDEGKTTISYGVVNLRLTKVSYDDQVSY-----873
QY 993 LENGVTDLTGIOKISRTHGRVTOIKDSSIKTTLNVDLNRHI-----GSQV--TDLA 1044
Db 874 -----NYDAVGNRLTKMDSHGTTAYKYDKLNRLLSVLNPDQCKVSYTYNK 918
QY 1045 TCHML-----TTTVEFDGLNREIGRKLCDSSGHTLD---IQOSWLKTCOLANRIVKL 1093
Db 919 VGNRVKMTYPPGKTTISYDAVNLIG--VIDSGHITSYAKNGNLKMTNTPNGV---973
QY 1094 NGVLRTEQYSDSRNRLNQYKDCGABCPDQYGHISIVTQNTFYDIYGNITACHTTPADG 1153
Db 974 -----KTE-YSYDKANRL-----VELINKNTTQVSSYKYVTLDAAGNRLKVDEQLAEG 1020
QY 1154 TED-----HATFKPANPDPCQLTEVHTHDPMDN--IRLKYDKAGRVINIT- 1199
Db 1021 VESGDSSELKESQLTTTGY--DKYRLTKV-----DYPNKTVSYKYDSMGNRISMTT 1072
QY 1200 --DNHGNTENFTYTLGRQNGQSVGYDPLNRLVSKQTDITLDCELY-YRETMVLNVEVR 1256
Db 1073 NVDGIGTISYKYDAADQLQSGNISYDKKGNLIKRVNSTQFMSYDE--ANRLK 1129
QY 1257 N-----GEMIRLLRT--GETTIAORASKVLLTGTDSQQSVILT 1293
Db 1130 NVSEFVNTPKSSYNFEDGDNRIKTTINGE-----NAQSTKYVLDINSALPQVLTE 1185
QY 1294 SDKQNL-----SOEA-----YSAYGKHKSTANDASIL-----1320
Db 1186 SDTKNTTCYVGTDLISMTSENAYEYHYHVDG:GVSRSLSDSGIIKNVTLYDAFGQVQK 1245
QY 1321 -----GYNGERADPVSGVTHLNGYRSYDPTLMRFHTPDSLS--PFGAGGINPYSY 1369
Db 1246 EIGTVDNNEFRFTGQMDDETGLIYLARY--YDPSVGRFITKDVIGRRVTTQOSINRYVY 1303
QY 1370 CLGDPINRSPSGH 1383
Db 1304 TTNPNVNLVLTGY 1317

RESULT 3
US-10-156-761-7834
; Sequence 7834, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7834
; LENGTH: 2364
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7834

Query Match 4.7%; Score 420.5; DB 14; Length 2364;
Best Local Similarity 21.6%; Pred. No. 1.8e-23;
Matches 320; Conservative 190; Mismatches 550; Indels 423; Gaps 73;

QY 134 RIITHKSGDIEVLTCFNNAFDLKVPKLLNPAGHAIYIDWNFEATOPRLNRIYDDLDGHD 193
Db 702 RKTGSGDDIVLPVINFNRDLD-----NKVGA---ELNF-----PRIKEIHGDL-----744
QY 194 IPLLNLLEYOGLIKITLTLFPQOKEGYRTELRLNQLNSIHNFSLNENPLTWSFCYTPI 253
Db 745 -----GATTVSYGFAN-----ACDIDHLPQAS-----NTQDCVQW-KWTPE 781
QY 254 GKNGILQWIT-----TSMTPAGGLKETVNYG--NNNQGHFFPQSANLPVLP-- 297
Db 782 GETDSKTGPKFKPLVTQVQVDFVTNODGAPVMTTSYTYEDGAGWHF---TNDPLIKDE 838
QY 298 -----YVTLMKQV-PGAGPAIQAESYVTSNHYVGGSGIWNKLDNLYGLMTRY 347
Db 839 DESWTDWRCYQEVQVTTGAGACQK-----TKSWLYRGLSGDRTSKAD-----881
QY 348 NYGSTESRRYKDKEGHDQIVRIERTYNNY-----HLLTSECKQONGYIOTTETAYYA 399
Db 882 -ASATKTVTVDDGQ-----NNYDSDLSGRILSTSLRDTG--TSHERTYHK 927
QY 400 IIGHN---PDSQPSQOLPKTKTETWRSADNSYRSEITETTTDES-----GNPLTKVI 449
Db 928 YNDHNTAQVGGLPDARFVREKETTTNTKVSQWREHTVETETEDDTGEGASTTTEGLEPMR---984
QY 450 KDKTKQKIISPSTHMEYPPAGEVDN-CPPEPYGTRFVKKIIQTPYSEFKDDPEKFTQ 508
Db 985 TDDWGQSSVS-----DNRC-----TTYGRAYNTD-----1008
QY 509 YEYSILGSGSHVTLKIERHYS-----ATOLLNSTLFOYN-----543
Db 1009 -NYDSTGAQRWTVLQDQVKHYSVGCSSIAQSKQGYASTLYDNATSVANKPVDGNPTES 1067
QY 544 -----TDKSELGRLLKQTE-----CTKGENGKTSYVVRHFTVTKQDDTLQO 584
Db 1068 RYTKSGSVRSWCSYDDAGRWMSDESHNRILTLYKSPANTWPLNGVIVITPPDGPALP 1127
QY 585 SHSIITHDNFTIHRQVRSRYTGRFLSDTDKDIIVTQMSYDKLGRLL-----631
Db 1128 AHTALTSTAWT-----SREWGKFTSIQDANGNVTKMSLDAAGRLVREVRPTETGSSPS 1180
QY 632 -----TRTLNSGTPYANTLTLYDYELNNLQDNRPPFVI 664
Db 1181 MKFSYTIPTSTNSAGVDAVDGYPHAVTHVLQSGTTLSSHAYVDGLGRAR-ETQTPMGN 1239
QY 665 TTTDVNGNQLRNEFDGAGRHYSQLKQSDGQKFTYTHIQOYDEQGR-----HH 713
Db 1240 GVDAAATGNEVEN-----RQVSVTRYDSAGN---VTGTSVAFRNQGTAGSGSPSAKVED 1290
QY 714 TSTYSIDYLTN--GRQOTDPDKVHLMSKSYDNWQCIANTHWSYGVSEKITVDPITLTKATK 771
Db 1291 LPSYITDLVDWAGRAITS-----RLQVNGASQDAGRVTFTYDGDFTSVKKNVDAADATVD 1346
QY 772 QLOQSNNNVOTGKEVT-----TYTPSQOPIQITLFDAGHLQSCHTLTRDGMWRVRET 826
Db 1347 VYGQVSKVVEHTGSATYVTAAYTKDELKIKIT--DPRGN-DTSYTYDWAQORTATDSD 1403
QY 827 AIGQCTIYQDYNRVIQITLPGD-TIVNRKYAPFSTDTLITDIRVANGISIGOOTFOLS 885
Db 1404 AGVST--EYDKNGQIKSITNSGGKTVLDYDAGLRKTA---VRSGDTLEAATWBDGLN 1458
QY 886 RLQSQDQGRVWAVTYAGNDQCPSTVITPDQGIHQYQD-----ELDDAVLQVASENI 940

Db 1459 -IT-----GGK-GQITSTGRDNGNTYTTKTGEF-DVRGRPLKTTVTIPDVTVKGLAGS-Y 1510
QY 941 TQOFSYNPVTGALLKAVAGQSLTPIYF-SGRKXENINDMKMSYLWTLRGLENG-Y-- 997
Db 1511 TTGFTYD-----AADHVSVAYPAGGLPAETVN-----AYDGYGR 1547
QY 998 -----TDLGTTQKISRDRTHGRVTQIK-----DSSIKTTLNYYDNLN--RHIGSOV 1040
Db 1548 PLRLQALQTYIRSTGYDAYDLSLTPRSYGVDSVLPFCIGAAQRTYSYDSDNGTRELKSA 1607
QY 1041 TDLATGMLTITVEFDGLNREGRKLCDSGGHGLDIQOSWLKTOQLANRIVKL-----NGV 1096
Db 1608 T-----TTT-----LNKVSERQKDT-----YTYDLAGKLTLEQASGQ 1642
QY 1097 LQTEQYSVDSNRNL-NQYK-----CDGAECPDKYGHISIVTQNFYDIYGNITACHT 1148
Db 1643 TAQOCFLYDQARLTNAYTHTTTGICADKTKTASDFKGTAPYQYATYDRLGNLQITN 1702
QY 1149 TFDAGT-----EDHATFKCANPTDPCQLTEV-HHTHPDMPDNIRLYKQAGKVI 1196
Db 1703 TDSAGTATLHXYLPGYDGTGTTTANANQPHGVKIDHKCTGSTITKSDQLYYFDDGTMK 1762
QY 1197 NITDNGNTENFTYDLGLQ-----NGQSV---YGYDP-----LNLRLVSKT-----DTL 1240
Db 1763 QRVE-FGITTDTVTWTPQGLEAVKTKSGSGSELTTRYAYDADGNILVTRTPQETVASIDGM 1821
QY 1241 DCELYVRETMNLNVEV-RNGEMIRLLRTGETIIAQORASKVLLTGTDSQOSVILTSRKON- 1298
Db 1822 ELATNGTIVTATRVASGTATVARTTEGITA--GKVTYLMADQASTOLAVDASTG 1879
QY 1299 -LSQEAYSAYGKHS--TANDASILYNGERADPVSGVTHLNGVRSYDPTLMRFTPDLS 1356
Db 1880 ASTREYTPFGERSGTLPTGDNGLGKTEDTSTGLSL--GARAYDNLNGLRFLSPDPL 1937
QY 1357 -SPFGAGGPNPYSYCLGDPINRSDPSGHLNMQAWTGIGMGIA 1398
Db 1938 ATPYAPQNLNAYATNPNISYSDPSG-LCRADICGDGYPVAG 1979

RESULT 4

US-10-156-761-7990
; Sequence 7990, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIOYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7990
; LENGTH: 820
; TYPE: PRI
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7990
Query Match 4.4%; Score 395; DB 14; Length 820;
Best Local Similarity 23.8%; Pred. No. 3.2e-22;
Matches 242; Conservative 120; Mismatches 375; Indels 280; Gaps 50;
QY 501 DPEKFIQVYSLIGSQSHVTLKIEERHYSATOLLNSTLFOYNTDKSELGRLLKQTEC-T 559
Db 26 DTEGKTTVRYDANGRVOKIT-----TASGRVTVFTYD-DANRVTSMLRGTFGNS 74

QY 560 KGENGKTYVWHKFTYTKQDDTLQOSSHITTHDNFTIHRQVRSRYTGRLFSDTDTKDIV 619
Db 75 DGHGTPTW-----TYAYTSDSVTAAGTTATDPTET-HATKYQHDSDGVQSDVTDAMGHK 127
QY 620 TOMSDYKLGRLLRTRT--LNSGTPYANVLTVDYEL--NMLQDNRPPFVITITDVGNOLEN 676
Db 128 RSTKEDANHSIDTSDAMSGSTTFCNVDTYGFNTRNLE-----TITQPTGGKTVN 178
QY 677 EPDG-AGRHVSQCLKSDGDGKFYTIHTQQYDEGRHHTSTYSYLTNQRQDTPDKVHL 735
Db 179 HQOTIAGADVPKDSNPDGKTDFT-----YDAVGN-----TW 211
QY 736 SMSKSYDNWGOIANHWSYGVSEKITYDPIITLTA---TKLOSNNSNNVONGKEVITYTPS 752
Db 212 SVAQTGTGGGNVSYTY-----NPASPTCGGFAGORCTEKTAKTAAKTVTT--- 256
QY 793 CQPIQITLFDAGHLQSCH-----TLTRDGDWRVRKETDAIGCTIYQYDNYN--RV 842
Db 257 -----TFHYDSAGNLDVTVPAPLAKITYTYDALGRKTKVTDARGVTVTYTYDNRDTRI 311
QY 843 IOTILP-----DGTIVNRKYAPESTDTLITIRVNGISILGQOFTDGLSRLT--QSQ 891
Db 312 VDTNKRARVEWYDGDGNLTOR--TDSGTIKYD-----FDFLQRETIRTLQ 356
QY 892 DQGR-VMAITYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEITQOFSYNPVT 950
Db 357 DGSQTLAY-----TPSGNVVYQ-----DPAG-----TVDYTNVEV- 398
QY 951 GALLKAVAEGOSLPIIYPSGRK--MENINDMKMSYMLTLGLENGYTLDTGTTOKIS 1008
Db 399 -----NKLAEKDPAGRVITYKYNNDVKTIT--TYFG-----GTQCKVD 426
QY 1009 RDTGRTVQIKDSSIKIT-----LNYDDLNRHIGSOV---TDLATGMLTITVEFDSL 1058
Db 427 PDSRSRPTIKTSPKGTVDLAYSQYGTSGTTEGSKIRSSDNTVG--LKTITYYDGA 484
QY 1059 NREIGRKLCDSSGHTLIDIOOSWLKTOQLANRIVKLNGVLORTQOYSDSNRLNQYKCDG 1118
Db 485 GRFSYAE--EKKGTLLN--SSWLYCYDLAGNL-----TSGTAAGCPGATTTVND 531
QY 1119 AECPTDKYGHISIVTQNFYDIYGNITACHTTTFADGTEDHATFKFANPTDPCQTEVHHHT 1178
Db 532 AQLTAKNGS--TTNWSYDKIGNETAGAST-PEGTRAEKW-----TDSQLTSL--- 578
QY 1179 PDMPDNIRLYKAKRVINITDNGNTENFTYTLG--RLONGQSVYGVDPNRLVSQK 1236
Db 579 -----TVGKK--TYTQYGVSTQDSERIRLGDYTFHNGPIGLAG----- 614
QY 1237 TDTLDCELYRETMLNVEVRNGEMIRLLRTGETIIAQORASKVLLTGTDSQOSVILTSK 1296
Db 615 TSTAGVDTGFNR-----EPGQTLNSMTTGKAYCYLTDALGSLVIALTDE 658
QY 1297 --QNLQSEAYSAYGKHSKSTANDASILGN--GERADPVSGVTHLNGVRSYDPTLMRFT 1352
Db 659 TGAKNVTYYSFRGVTASTSEKNPQYRFPAGYQD--VTGLIYHARY--YDNIORFNS 715
QY 1353 PDSLSPFAGGGINPYSYCLGDPINRSDPSGHLNMQAWTGIMGIAGLLITITATGGMAIAA 1412
Db 716 PD---PSQOE- NPYLYAEGDPVNRIDPNGLFSDALDTGSDIFGVVTVGCLAGVTAAG 771
QY 1413 AGGTAAATASTTALAFGALSALTSITSDITSVSGALEDASPASSILGWSMGMGAAG 1469
Db 772 TGSIAAAA-----VGVVGVAGVSGVGLAVVGVSCALG-GAAG 810

RESULT 5

US-10-156-761-7572
; Sequence 7572, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO

```

; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7572
; LENGTH: 1250
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7572

Query Match
Best Local Similarity 4.4%; Score 392; DB 14; Length 1250;
Matches 271; Conservative 187; Mismatches 452; Indels 368; Gaps 66;

QY 254 GKGILGOWITSMTPAGGLKETVYNNQGHFPQ-----SANLPVLPVYVTLMKQVPG 307
DB 59 GMEGVAEGGLRK-TARNAVEDAEDASQKAKICEKDPVDVATGRVLPQI-----DVSL 113
QY 308 AQPAIAQAEYSYTSYNNYGGGSGNGIWNKLNLY-----GLMTEYNYGSTES 354
DB 114 PGQLPLVVKRGFESSYRLGGWFGPTWSLTDORLEVDTAGVVLVGEDGLVLAIPH-----PA 170
QY 355 RRYKDKEGHDIQVIERIYNNYHLLTSECKQNGQVIOETEPAYVIAIIGHNFDSPQSFOL 414
DB 171 PGVPTLPSHGPRWLDRTDGGYTL-----TDQKSGHIR-----203
QY 415 PKTETETWRSADNSVRSIETTTTDESQNPITKVIKQKTKQIISPSTHWEYVPPAGEVD 474
DB 204 -----HFVDRSDTLAVLEQLDDRNENWIT-----FEY-----DAD 233
QY 475 NCP-----PEPYGFTREX-----KIIQTPYDSEFKDDPEK-FIQRYSLIGSQSHVTLKIEE 526
DB 234 GAPRYLTHSGGYRLRISTEAGRVTAHLAAGVGGQDLIRGY-----IDGHL-----283
QY 527 RHYSATOLLNST--LFQYNTDKSELGLLKQTE-----CTKGENKGYTSV 569
DB 284 -----TEVNSSGRLQVQCD--ELGRITSWTDNDSHFSYAYDDEDRCTH-QSAGHL 335
QY 570 VKETTYTKQD-----DTLQQSHSIITHDNFTIHRQVRSRYTGRLFSDTDKDIVTOM 622
DB 336 RSTFAYGAVDPGTGAHTTTVDSYGQTH--YLNR-----RCQVIAETDALGAVTRY 386
QY 523 SYDKLGLLTLRLTNSGTPYANTLYDY-ELNNLODDNRPPFVITTTDVNGNLQNRNBFDA 681
DB 387 QDRYNNELLSQT-----DPLGHTTSFRYDDAGNLVAATRP-----DGREARAYNAL 433
QY 682 GRHVSQCLKDSGDGKPYTHIOQYDQGRHHTSYSDYLTLNGRQOTDPPDKVHLSMSKY 741
DB 434 GLPVK--LVNPDG-----TITRQTFDERG-----NLTSTVDS--GQTRFGY 472
QY 742 DNMGQIANTHWSGVSKITVD-----PILTL-----ATKQLQSNNNVQTKVITYTPSQ 793
DB 473 DEGRSLTSMTDPLGHTTGVCDRAGLPLTVTDPLGAVTRYERDAFGRTTA--ITDPTGAT 530
QY 794 QPIQITLFDAGHLQ-----SCHTLTRDGRVREKETDAICQCTIYQYDYNVRVIQT 846
DB 531 TLEWTV--EGHLSRTAPDGSSEWYDGENCTSHTDPPVGGVSLFEYTHFDLPTART 587
QY 847 LPDGTIVNRKYPFSTDTLITDIRVNGISLGQOTFDGLSRLTOSQDGGGRVAVTYVTSAGND 906
DB 588 GPDGV-----RY-EFEHD--TELK-----LSQVTPNH--GLTWNAYVDAERG 624
QY 907 QCSTWITPDGQIHYQYQPELDDAVLOVAS--NEITQQFSY-NPVTGALLKAVAFGQSL 963

625 LVAETDF--DNRTLTVEY-----DFAGRLASRTNALGQMTAFERNELGQIVRKDAAGQAT 677
964 TPIYVPSGRLLKVENINDMKMSYLVTLRGLENGVTDLTGTIQKISRDTGRVTVQIKDSSI 1023
678 TYAYDFDQLAQ-----TGPDGTALTILDRHGRMISEAVNGR 716
1024 KTLINVDLNRHI-----GSQVT-----DLA-----TGHMLTTTTFDGLNRIRGRKLCD 1068
717 ALTVDYDELGRTRTRTPSSATTIWSYDVAGRETSMTGSGRSIDFVYDEAGRELGRPF-- 774
1069 SSGHTLDIQOSWLKTOQLANRIVKLANGVLQRTQSYSDSHNRLNQKCGAECPTDKYGH 1128
775 --GH--TLEHSFDALGELTSQSV-LGPAGRTHURQY-----TTRAD-----GH 814
1129 SIVTON-----FTVDIYGNIPACHT-----TF-----ADGTEDHATFKFANP-----1165
815 LIGIEDQLSGKRFFDLGRVTAHAVNVNTETIYDAYVGNQTSAGSWPAGHPQOEALGNR 874
1166 ---TDPCLTEVHTHPDM-----PDRIRLYKDYKAGRVINIDNHNGTEN 1207
875 TYTGTSLTRAGQVRYEHDGLRIILKQTRLSKRPDTWRYTDAEDRLTSVTPDGTWR 934
1208 FTYDTLGR-----LQNGQGSYGYDPLNRLVSKQTDLDCELYYRETMVNEVR---N 1257
935 YTYDPLGRRTAKRLAEDGETLV-----ERVDFTWDGTVLCEQTTTSPDLPHQVTLTW 988
1258 GEMIRLLRTGCTIIAQORASKVL-----LQTDSCQSVILTSQKNSLQSEAYSAY 1307
989 HQGLRVPVOTERIVAADAPQOEIDSRFFAIVTDLVGTPSE---LLDERGEIVMRTRATLW 1045
1308 GKHKSTANDASI--LYNGERADPVSGVTHLNGVRSYDPTLMRPHTPDLSLSPFGAG-GI 1364
1046 GSTTWAKNSTAYTPLRFPQYVDPETGLHY--NYFRHYDPEIARYVTPD---PLGLGPAP 1100
1365 NPYSYCLGDPINRSDPSG 1382
1101 NPAAV-VPNPHMWADPLG 1117

RESULT 6
US-10-156-761-7751
; Sequence 7751, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7751
; LENGTH: 2386
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7751

Query Match
Best Local Similarity 4.4%; Score 391.5; DB 14; Length 2386;
Matches 379; Conservative 214; Mismatches 688; Indels 601; Gaps 93;
163 NP---AGHAIYIDWNFEATQPLNRIYDLDGHDIPLNLEYQGLIKTILTLFFPQKEGY 219
```

Db 698 NPDGTTCKTLMLDYIQRKT-----YGDGDDVLVPVINFRTDL-----DNKVG- 740
Qy 220 RTBLRFLNQLSINHFSIGNENPLTWSPG-----YTPIGKNG 257
Db 741 SABLNF--PRIQIHG-DLGATTKYSYGFANACDIDLHFTQASNTQDCYWKQWPEGETE 797
Qy 258 ILQOMITS-MTAPGGIKETVYNNQHHPPQSANLPLVPVILMKQVPG-----AGQ 310
Db 798 SKTGFKKFLVTKVEVDPIV--TTWQDG-----APVMTISYIEGAGWRITGD 844
Qy 311 PAIOAE-----YSYTSNHYVGGSGNGIWNKLDNLYGL-----MTEYNYGSTESRYKD 359
Db 845 PLVKDESDSWDRGQYQETVTTGADTGQTKKYLWYRGLDGDRTSKTDTSATKTVTND 904
Qy 360 KEGHDQIVRIERTYNN-----HLLTSECKQNGVIQTTETAYVALIGHN---FDSQ 408
Db 905 GDG-----NNYADHANLAGHTLSTSLRDDTDI--SHERTYHYKSHNTAQYDGL 951
Qy 409 PSOFQPKTKETWRADNSYSEITETTFDES-----GNPLTKVKKKTKOKILSPS 461
Db 952 PDAHFVRESKSTNTKISGMRHVEHVEYDDSEASTTFLGPMR-----TDDWGOSNYS-- 1006
Qy 462 THWEYPPAGEVDN--CPPEPYGTFRVKKLIQTTPYDSEFKDDBKFIQYRYSLLIGSQSHV 520
Db 1007 -----DNRC-----TTGRAYNTD-----NYDSTGAQRWT 1031
Qy 521 TLKIBERHVSATQLLNSTLFOVNTD--KSELGRLLKQTECKTGNGKTYSVVHKFTTKQ 578
Db 1032 VVODQVKHYSVG-----CSSIADSNQDGYTSTLYDNATSIDANKPVDGNATEV---RTYTK- 1084
Qy 579 DDTLQSHSITTHDNFTHRSQVRGRYTCRLFSDDTDKDIQVMSYDKLGRLLTERTLSNG 638
Db 1085 -----AGYRSWTG-----YDKAGSIWVS--EDG 1107
Qy 639 TYPANLTLYD-----YELNMLQDN-----RPPFVI-----TTTDVNGNQ 673
Db 1108 XHNRSTITYSPANTWPMNGITSPDPDGTATARGPLTFTEWTSRFPWATPITTSKDANGNI 1167
Qy 674 LRNEFDGAGRHSQCLKDSGDG---KF-YTHITQO-----YDEGRHHTSTYS 718
Db 1168 TKVTLDAAGPEVEMKPIETGSSPSMKFSYIPISTNSAGVPSDAGPYVASHTLQSGS 1227
Qy 719 DYLTN-----GR-----QOTDPKVLHMSKS-----YDNMGQIA----- 748
Db 1228 TYLVSHAYTDGLGRARETQTPLPSPVDPATQIVFFROVAVTRYDSAGOVGTGASAVFRNQ 1287
Qy 749 NTHWSYGVG-----EKITVDPITLTATKQLQSNNSNVQCKEVTTY-----TP 791
Db 1288 GTAGSGGPPSPQSDLPYSYDVLVDWAGRTVSSQIQVKGTQKAGRVDTSYLGDYTSVTP 1347
Qy 792 SQOPIQITLDEAG-----HLSCHLTIRDGW---DRVKETDAIGQCTIYQYDNYR 841
Db 1348 VDSATDITDYVGVQSVKVEHTASAYTTAYGTAKGELAQLTDPRGNNTLYTDWAAQ 1407
Qy 842 VIQITLPDGTIVNRKY---APESTDTLITDIRVNGISLGQOTFDGLSRLTQSQDGG---R 895
Db 1408 RKTDDPDAGLSSESEVNGQVSQITATNDVQVTLTVG---YDNLBRATSVRSAGADELA 1464
Qy 896 VNAY-----TYSAGNDQCPSTVITPDQGFHYQVQPELDDAVLQVANSNETQQ 943
Db 1465 AMWDDPPAATGKGQITSAVSADASGNTYTTKTGK-----DERGRPLNTTVTL 1514
Qy 944 FSNVPTGALLKAVA--EQQSUTPIYP--SGRLKMNENI-----NDMKKMSYLVTLRLGLENG 996
Db 1515 TTVNGLAGDYTSVTVDAADHITSVSPAAKLAALKVTTYTYDDYQOPTRLTSLG--- 1570
Qy 997 YDITLCTIQKISDTHGRVQTKDSSIKITLN-----YDNLNHHIGSQ-VTDLATG 1046
Db 1571 ---GTAYDNTTYDAYGLVE--RDYGAFFGGNGQAORQYGYDSN---GTRMLRSIAT- 1622
Qy 1047 HMLTTTVEPDGNRIGRKLCDSSGHTLTDIQCSWLKTKQOLANRIVKLVGLQRTQYSYD 1106
Db 1623 ---TTINDLVSEAQKDTVLYDNTCKLTRE------QASGQTAQSQCLRYD 1666

Qy 1107 SHNRLN-QYK-----CDGAECPTDKYGHISIVTONETYDIYGNITACHTTFADGT--- 1154
Db 1667 DOSRLUTLAYTHHTAGNCADTTKTSDFKGTSPYQGTGYTDRLGNLQSVTDTNSAGAATTR 1726
Qy 1155 -----BDHATTFKFPANPDPQCLTEVHTHPMPDNPRLKYDK-----AGRVINIT 1199
Db 1727 DYLYPGYDAGTWTVAN-----ADQPHGVR-KINKVSAGTTTAAAGTYTYA 1771
Qy 1200 D-----NHGNTENFYDYLGLR-----QNGQSVYGYDPLNELVYSQKTDLTDLCEL 1244
Db 1772 DGAMKORVEGTTDYTWISRLGLRATVKTWTSGSDLTRTYTDASGNLLVTRTP----- 1825
Qy 1245 YVRETMV---NEVRNGEMIRLLRT-----GETIIAQORA-----SKVLLTGTDSQOS 1289
Db 1826 --QETVASIGGTETLRTTDDIGISATATRYVSPGATTVAMRTTDDGNTVNGKITVLMGDTQAS 1883
Qy 1290 VILTSQKQ--NLSQEAYSAYGKHS--TANDASILGYNGERADPVSGVTHLNGCYRSDPT 1346
Db 1884 TOIADVAATGATATRRYTPFGDERSGSLPTGTNNHFLGKTEDNTGLSLL--GARYDPS 1941
Qy 1347 LMRHTPDSLS--PFGAGGINPYCYCLGDPINRSDPSPGHLWSQA-----WTGICMGIA 1397
Db 1942 LGRPLSPDPLSTPYDPQNLAYSYSNGNDPINYSDPSGLIKLNSDGTQCSGDKGCGFTV 2001
Qy 1398 G-----LL-----LTATGMAIAAAG-GIAA 1418
Db 2002 GGGGDTPTPAPTVQOQLVDLLPREENGWDADRLAQVWHYQITQGGYWDAPVGGDRT 2061
Qy 1419 ATASTTTL--AFGALSVTSDITSIV-----SGALEDASPKASSILG 1459
Db 2062 GWACFGRTACSEAFVWKETHDPAKAKRVAATFCVENPKCGADNGAYDSMKSEASAVPI 2121
Qy 1460 VWSMGWGAAGLAEBAIKGTTKLATHLCAFAEDGENALLK--STSESRILKWG-----V 1510
Db 2122 LLAGEMGAA-----FSKVLKARGCSFKPTTRVLKMDGKTKPLGKIPGDLVEAADP 2172
Qy 1511 TRSLDREI-----VRNEEGQVQIKDHS-----RGYTDNFMKGEOAILV 1548
Db 2173 TSGHREVREVTAVHLNDDDLVDLSIRGLDRIQTLLHTTARHRIWDDTAQVWEQAGRLI 2232
Qy 1549 HGDYDGLYHTTEGKH-----NCKGPTYRTEPEQLVDYLDKNNIVDLTQGGDKP-- 1597
Db 2233 TGHR-----VNTSGNQHATITSVLAQRAADMYDLTVEGLTY-----YVLAGEIPVL 2280
Qy 1598 VHLISCYKSS-----GAADKMAKVINRPIVAYSNNKPTISOGIARIERKDFPKSTYHS 1651
Db 2281 VENGSCWSSTNRKTSVKNAFGHWKHH-----KSBFPNLNNAKEYVEAGTDFLRST--- 2330
Qy 1652 YDPKILIGRTEKTVKPKTRP 1673
Db 2331 -DPS--VLTRFRANGDVIRNP 2349

RESULT 7

US-10-282-122A-46365
; Sequence 46565, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Chisen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 46565
 ; LENGTH: 2234
 ; TYPE: PRT
 ; ORGANISM: Bacillus anthracis
 US-10-282-122A-46565

Query Match 4.3%; Score 386; DB 12; Length 2234;
 Best Local Similarity 18.7%; Pred. No. 1e-20;
 Matches 339; Conservative 236; Mismatches 633; Indels 608; Gaps 76;

```

Qy 21 FTQANNFSAVSGVDPR-----TGLYNIQITLGHVIG-----NGNL-----57
Db 748 YSKVGHAESEISHFVPRFSODSGFIGNFDYWASIPVLNGKVNATNGNFMSEKDITLSG 807
Qy 58 -GPTLPPLTSLSPLNKTDIGFGIFNFGLSV---YDRKNLSLSL-----TG-EN-102
Db 808 RGPDSVVERTYNSQSKVGLFGTCWSSGLEERWADGNLILLISTDGANITRTGDNK 867
Qy 103 -----YKVIETDKTVKLOQKLDNIRFEKDKLKENCYRIHKSQDIEVL 145
Db 868 YQAPTGYILEIKOVSGGYEIKDKQTV-----TFYKSGDAQR 905
Qy 146 TGFNNNAFDLKP-----KLLNPAGHAIYIDWNFEATQPLNRIYDDLDGHDPL 196
Db 906 IEYTKDKYGNITTYEYDGAERLSKVKNASGKELVL--QYDGNKKAAVLI-----GPDNKT 959
Qy 197 LNLEYQGLIKTILTLFPQ--KEGYRTFLRNLQNSIHNFSLGNENPLTWSFGY-----250
Db 960 ITFNYDGLLVSSITPFGKVKGYD-----NGVLTSIYDPQHTDAKPKYKTSYAYENDR 1013
Qy 251 -----TPIKNGNL-----GOWISMTAPAGLKETVYNNQGHHPQSANLVLPLYVTL 301
Db 1014 LVKVTPLKATTLAVNNGSKVLTNPKGRKTVTYND-----AGNFV-----1057
Qy 302 MKQVPGAGQPAIOAEYSYTHNYYVG-----GSGNIWNKLDNLYGLMPEYN- 348
Db 1058 -KTVEDVGRNLNLTTSYEYNNANLVKITPKNQETATYDGNVTVDEMTEKEFYNK 1116
Qy 349 -YGSTESERRYKDXE-----GHDQIVRIERTVNNVHLLTSECKQNGYIQTETAYAYAI 400
Db 1117 DNGIILKATDNEDRKTTVAYVGANTEVSQTDQGAN-----TSSVIHHDQYCNPIETSKELS 1171
Qy 401 IGHNFDSQSOFOPLPKTKTETWSADNSYSEITETTFDESGNPLKVIKDKTKQKILSP 460
Db 1172 AGGNLQNPFS-FEM--NGIEKWVKVDNNSGSI-----KDATPACGLGGESSLKITTK 1223
Qy 461 STH--WEYPPAGEVDCNCPPEPYGFTFRVKKII-----QTPVDSEFKDDP 503

```

```

Db 1224 ATNNDWGYIAAIOEVLTEPNTTYTLSCMVKTDLVNGAAFFNVQSLNENGAGIDGWHDR 1283
Qy 504 EKFIQYRSLIGSQSHVTLKIEERHYSATOLLNSTLFOYNTDKSELGRU--LKQTECTKGE 562
Db 1284 HNKVQGTSDWVRQ--VFYKTEQ-----TRKVIYLOVENGGATSGSAFWDKIQLEKGE 1337
Qy 563 NGKTSYVVKHFTYTKQ--DDTLQO-SHSIITH-----DNFTIHRQSVRSRYTGLRFS 611
Db 1338 VSSFPNVLNSSFENWPGFVQWVRSCSQHCERNVDVSDSFTGHSSIWER-----S 1391
Qy 612 DTDKDIVQMSYDKLGRLLTRTLNSGTAYANTLTVDYELNLDNODNRPFVITTTDVG 671
Db 1392 EYCPNDI-----GYNRVILNOKKAEVTVLTAMSKSENVVD-----1428
Qy 672 NOLNEBFDGAGRHVSQCLXSDGDGKFYTIHTQQYDEQGRHHTTYSYDLTNGRQOTDP- 730
Db 1429 -----APDKLSKDYAVLAETYYQDG-----TVVNYTSPFSTNDW 1464
Qy 731 -----DKVHLSMSKSYDNWQOIANTHWSYGVSEKITVDPITLTATKQLOSNS 777
Db 1465 NRSAAVIPAKKPIQKIEIFLLFRKNKG-----KVFDDIILLEGNALIKNE 1511
Qy 778 NNVTGKEVTYTPSQOPIQITLFDAGHLQSCHTLTRDGDWRVRKETDAIGQCTIYQD 837
Db 1512 YD-NDGNVVAITY-----DEEGQK---NFTYDASGNKKSETDEKNTKLYDYN 1555
Qy 838 NYNRVIOITLPDGTIVNRKYAFSTDTLITDIRVNGISILGQOTFDGLSELTSQDGRVW 897
Db 1556 KNLATKVTLKNGTSVNYRYD-----VTS-AT-----DSRGGKTD 1702
Qy 898 AVTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEITQOFSYNPVTGALLKAV 957
Db 1577 -----HNGNTEKSVMPFGKTCHEKYVDVNDKNTYVIDALNRRIEN--TYDENANKIKTKM 1631
Qy 958 AEGQSLTPIYPSGRLKVENINDMKMSYLTWLRGLENGYTDLTGTIQXISRDTHGRVTQ 1017
Db 1632 PNGSILESVDYADRVAVGEK-----RNGKDSFT-----FERDQNGQVTK 1670
Qy 1018 IKD--SSIKTLNLDLNRHIGSOVTDLATGHMLTTFVEFDGLNREIGRKLCDSSGHTLD 1075
Db 1671 VKDLVNGVERTYTKADR-----VTS-AT-----DSRGGKTD 1702
Qy 1076 IQCSWL-----KTQOLANRIVKLVGLQRTEQSYDSRNRLNQ--YKCDGAECPYTKY 1126
Db 1703 ---WAYHDKANSKTEKLEQVTVQGY--TNKVSVD-YNTLDQNRVTDGSO-----1748
Qy 1127 GHSIVTQNTYDIYGNITACHTTPADGTEDHATFKF--ANP-----TDPQLTEVHHT 1177
Db 1749 -----TYRFDYDDQGNV---RTYTAGNGSGSTFNVDQANKIKOLVVGTSNLSLLSERYE 1799
Qy 1178 HPDMFPDNLRLKYDKAG-----RVINITDNHNTENFTYDTLGR-----LQ 1217
Db 1800 YDQSGNRRTKIKHEGAGKVTETNFVYDPIQLLNEVLNPGTTSYTYDGFGRNRSVKVIE 1859
Qy 1218 NGQGS---VYGVDPNLRVSKOTDPLDCEL-----1244
Db 1860 NGKETKTAATFNEGNQLVKFGNESLTYDVANGNRTSDGKYKYTWNEDDQIVAITKQGENN 1919
Qy 1245 -----YVRETMVNEVRNGEMIRLLRTGETI-----IAQQ 1274
Db 1920 APATKYDEDNRRTEKXVNGVQVTRFYDGDSDINPLIYETDNGTVLRQVYVSADGARLAWK 1979
Qy 1275 RASKVLLTGTDSQOSVI--LTSKONLSQEAYSAYGK-----HKSTANDASILYNGERA 1327
Db 1980 AQGQTLVYHVNPRGDVAVANTNQDEKAVATYEDAWGNVLTSDTKGIAAD-NPFYVAGVWY 2038
Qy 1328 DPVSQVTHLNGYSYSDPTLMRFHTPD-----SLSPFCAG-----GINPYCYCLGDPNIR 1377
Db 2039 DKEIGMYL-----IARYNPEHGVFLSVDDPDPGDEDDPVTMNGYTYADNPNVW 2088
Qy 1378 SDPSGHLSQAWTIGMGIAGILLATATGGMAIAAGGIAAAIAASTSTALAFGALSPTS 1437
Db 2089 TDPDGKAW-----LVPVVIAGAWAARFGAKYAI-----RYGAKYKKAOKS 2131

```


QY 1438 DITSIVSGALEDASPRASSILGWVSMGMAAGLAESAIAKGGTKLATHLCAFAEDENALL 1497
 Db 2132 -----GWY-----GKVAKSGWKNKSIQAQIPRIHKVGR---I 2163
 QY 1498 KSTSESRIRKMGVTRSLDRIVRNEERGQVVKHSHSGYDTNFMKGGEQAILVHGDKDGFY 1557
 Db 2164 KGDNDKGGYWGVIYTTK-----TKGKRTYS-----SFEFHTPHNGHY 2203
 QY 1558 HTECNKNG-KGPYTR 1572
 Db 2204 HLQNKYKYGKWR 2219

RESULT 8
 US-10-418-861B-55
 ; Sequence 55, Application US/10418861B
 ; Publication No. US20040010131A1
 ; GENERAL INFORMATION:
 ; APPLICANT: da Silva, Ana Claudia Raseira
 ; APPLICANT: Parah, Shaker Chuck
 ; APPLICANT: Quaggio, Ronaldo Bento
 ; APPLICANT: Reinach, Fernando de Castro
 ; APPLICANT: Ferro, Jesus Aparecido
 ; APPLICANT: De Oliveira, Julio Cezar Franco
 ; APPLICANT: De Laia, Marcelo Luiz
 ; APPLICANT: Setubal Joao C.
 ; APPLICANT: Furlan, Luiz Roberto
 ; TITLE OF INVENTION: Isolated Xanthomonas nucleic acid molecules, proteins encoded the
 ; FILE OF INVENTION: uses thereof
 ; FILE REFERENCE: FAPESP 205.1 US
 ; CURRENT APPLICATION NUMBER: US/10/418,861B
 ; CURRENT FILING DATE: 2003-04-17
 ; PRIOR APPLICATION NUMBER: US 60/374,620
 ; PRIOR FILING DATE: 2002-04-22
 ; NUMBER OF SEQ ID NOS: 85
 ; SEQ ID NO 55
 ; LENGTH: 1510
 ; TYPE: PRT
 ; ORGANISM: Xanthomonas
 ; FEATURE:
 US-10-418-861B-55

Query Match 4.1%; Score 361; DB 15; Length 1510;
 Best Local Similarity 21.2%; Pred. No. 5.4e-19;
 Matches 304; Conservative 158; Mismatches 466; Indels 504; Gaps 75;

QY 144 VLTGFNNAFDLK-----VPKLLNPAGHAIYIDWNFEATQPRINRIYDLDG 191
 Db 250 ILTRPNGNFYTKKSANGTWTDPDVRTELSEVQDNGTLTGQVVTATDSDKEQP--DLDG 307
 QY 192 HDIPLNLEBYQGLIKTLTLFPQGEYRTELRLNRLNLSIHN-----F 236
 Db 308 ---KLTGISY-----TDGQLTLTVAGRLQQLQVTTDTRGRRLPAYQADRI 350
 QY 237 SLGNENPLTWSFGYTPIGKNGILGWITSMAPQGLKETV---NYSNNQGHFPQS--- 290
 Db 351 QVGLPDGMVLAIGYDSQARL-----RSVTLTAAGAVASAIAGYDYG--ARFPDALTT 402
 QY 291 -----ANLPVLPYVTLMLKQVPCQAPAI-QAEYSY-----TSNYYGG---G 330
 Db 403 HRDEQOVYASVYDAQQRVRSVHGDTGKIDEATYASNTSTVSNALGNVTRTGIS 462
 QY 331 GIWNNKLDNLYGLMTEYNYGSTESRR-----YKQEGHQIVRIERTYNNHLLTSECKQ 385
 Db 463 KLGAQVAVAVQGLCEPTVGAFKRSYDNGYPPQVDEVDGAVATDWRNRLGLASKIEA 522
 QY 386 QN---GYIOTTTAYVAILIHNFDSPQSQFOLPKTKETWESAD-----NSYRSE 432
 Db 523 DNSSGGOKETLQT-----DWHPS-FRVP-TDQRTYDASDILVARTSWIYNSEGOA 570
 QY 433 ITETTFDESNGPLTKVKKDKTKTKIISSTHWEYYPAGEVDNCPPEYGFTRPVKKIIQ 492

Db 571 LTVSRTPSGGP-TRV-----TTORYCEDSD-----TAAG-----NC----- 601
 QY 493 TYPDSEFKDDPEKFIQYRYSILIGSQSHVTLKIEERHYSATOLLNSTLFOYNTDKSELGRL 552
 Db 602 -----SLPG-----LLASDGAFTDIADTSYTY-----YL 627
 QY 553 LKQTECT-----KGNGKTYSVVHKFT-YTKQDD-----TLQOSHSHITTHDNFTIH 597
 Db 628 ADDASCTASASTPCPRKGRKWKVINALCQTTLEYLAYDAGRPLSIKDTNGIVT--DYTYH 685
 QY 598 -----RSQVRSRYTGRLFSDTKDVTOMSYDKLGLLTRLTNSGTPPYANTLTVDYEL 651
 Db 686 PRGWLTKSVRG-----ADASSEADRIIDYVPTG--LVRQVTPDGAFTFTVD-AA 737
 QY 652 NNLDQDNRPPFVITTTVDVNGQLRNEFDGAGRHVSQCLKDSGDGKPYTTHTOYDEQGR 711
 Db 738 HRLTD-----ITDNAGNTVHTLNDAGNRVKEDTKDAAATLK--RTLRYVYNQLGQ 786
 QY 712 HHT--STYSD-----YLTNGRQQTDPKAVHLSMSKSYDNWGOI-----ANTHS 753
 Db 787 LKQATAASDPTDFAYDANGNAKVTDALATATQSEYDPLNRLSHLTQDVAGIKATKFA 846
 QY 754 YGVSEKIT--VDPITLTATKQLOSNNVQ-----TKGEVTVTPSQOPIQITLFDGAGH 806
 Db 847 YDALDNLTKVTDPKGLDITDYNGFGLVKLTSPDTG--VTSYT-----YDSAGN 894
 QY 807 LOSCHTLTRDGDWRVKETDAIGCTIYOYDNVNRVQITLDPDGTIVNRKYVAFSTDTLI 866
 Db 895 RAT-----QTDARGNTYAYSDALNRLTKVITYPS-----SLDVTY 930
 QY 867 T-DIRVNGISLQOQTFDGLSELTSQDGRVWAYTSAGNDQCFSTVITPDGOFIHYQYQ 925
 Db 931 TYDVTQTACTSG-ETF-SIGRLTKVQGGGAIQYCNRFCD-LVRKVQTSNGTALVRYD 987
 QY 926 PEL-----DDAVLQVASNE--ITQOFSYNPVTGALLKAVAGQSULTPIYPSGRL 973
 Db 988 YTVGGQLRRMYPDGAVVDYVRNAQGTTOGVTPAGGS--ROVLGNA---TYFFGPA 1042
 QY 974 KMENINDMKKSYLWLT-----RGLNGYVTLTGITQIKISRDTGHGKVTQIKDS 1021
 Db 1043 AGWTYNGRTLARQVLDYRPAIQDTRPGGLDVG-----GFDPAGNLTALTEA 1092
 QY 1022 SIKTT---LNVDDLNRHIGSOVTDLATGHMLTTTVRFDGLNREIGRLKLCDSCHTLDIQ 1078
 Db 1093 GNTPEISGLYDALGRITG--LTDGVITGVI-----DG----- 1123
 QY 1079 SWLKTQQLANRIVKINGVLQRTQYSDSR-NRLNOKYCDGABCPDKYGHSHIVTQNTFY 1137
 Db 1124 -----YSYDATGNRL-----STKVG--TATQIYTY 1146
 QY 1138 DIYGNITACHTTFADGTEDHATPKFANPTDPCOLTEVHTHPDMPDNIRLKYDKACRVIN 1197
 Db 1147 -----PNDSHLSAV-----AGVARIYDATG---N 1168
 QY 1198 ITDNHGNENTYTLQRLONG--GSVYGDYDPLNRLVSOXTDLOCELYYRETMJLVNEV 1255
 Db 1169 TTAIGGTARQVYDTSGRMTQARRAGAV-----TMNRYN 1203
 QY 1256 RNGEMI-RLLTGTETIIAQQRASKVL---LTGTDSSQSVIL----- 1292
 Db 1204 GKGEQVRFGLTNTYTLFDEAGHNLGVDYDSNGAPKQQAIIWLDDLPVGLLANANKLHYIE 1263
 QY 1293 -----TSD-----KQNLQSAQYSAKHKSTAND--ASILG--YNGERADPV 1331
 Db 1264 PDHLGSPRVVIDPTDPRVAVVMTSLKGEAFNGTAPNPDGDDGAAAVLDMRFFGQRFDAAS 1323
 QY 1332 GVTHLNGYRSYDPTLRFHTPDSLSFGF-AGGINPYSYCLGDPINRSDPSG 1382
 Db 1324 GLNQ--NYFRDYEAATGRYQSD---PIGIEGGISTYSYLSSEPVKYIDVLG 1370

```
; Sequence 76214, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 76214
; LENGTH: 843
; TYPE: PRF
; ORGANISM: Salmonella typhi
; US-10-282-122A-76214

Query Match          4.0%; Score 353.5; DB 12; Length 843;
Best Local Similarity 22.8%; Pred. No. 7.8e-19;
Matches 223; Conservative 121; Mismatches 308; Indels 327; Gaps 48;

QY 542 YNTDKSELGR-----LLKQTECTGKNGKTYSVVHKFTYTKQDDTLQOQSHSITHDNTI 597
DB 32 HHTDSGEQYRLDNWLAERSLCVTDMSMGT-----RCHWDAQGLVTAY 74
QY 598 RSOVRSRYTGR-----LFSDDTDKDIVTQMSYDKLGRLLTTLNLSGTPYANTLYDYE 650
DB 75 RDEAGQMTTFRWSDEERLLIGMTDAQGGKRYVYVDRGLHL-----TETHD-P 120
QY 651 LNNLQDNRPFP-----VITTTDVNGNQLRNEFD-----GAGRHYVQCL 689
DB 121 LGRVEQQTQHPVWHQPETEYDAAGVAMRYEYDERGNLQAVSDPLHQRTVYGYDRH-QQV 179
QY 690 K--DSDGDGKFFYTHTCQYDEQ-----RH-----HTSTYSDYLTNGRQQTDPKVLHLS 738
DB 180 RITDARGDKYL-----QWEDGQLMRHTDCSSQTAWFYDERLRLERVTDAES--NSTR 232
QY 739 KSYDNWQCIANTHSYGVSEKITVDPIITLTKQLOSNNSNVQTKREVITYTSPQOPIQI 798
DB 233 YSYDGNCHLTVMPFADGRTERYQPD-----AAGRLVKYTSFAG--QI 272
QY 799 TLFEAGHLQSCHTLRFDGDRVRKETAIGQCTIYQYDNYNRYVIQITLPDGTIVNRKYA 858
```

```
DB 273 TEWQ-----RDQGRVREQTATGRTAYEYDAYGRLTTLNNGESYRFRY- 319
QY 859 PFSTDTLITDIRVNGISLGQQTFDGLSRLTSQD--GGRVWAYTYSAGNDQCP----- 909
DB 320 -----DVLDRVTETQTFDGGRRRAYGYNALNAVTAIVYGGERG 356
QY 910 -----STVITPDQFIHYOPELDDAVLQV-----ASNEITQOF 944
DB 357 GEIRHGLERDAAGRLTAKITPE---TRTEYRYDAADRLLLEIRRRHRDAAEGEPEVI-RF 412
QY 945 SYNPTVGTALLKAVAEG-----QSLTPIYPSGRLLKMNINDKMKSVMYLTURGLN 995
DB 413 SYDSAGNLLSEETAQVLQHRVDVQCNRTEQMPDGR-----TLRYLYY 456
QY 996 GYTDLTGTIQTISRDTGHRVTQIKDSSIKTTLNVDLNRHIGSQVTDLTATGHMLTTTVEF 1055
DB 457 G-----SGHLOQI-----NLGRDVISEFTR----- 476
QY 1056 DGLNREIGRKLCDSSGHTLIDIQQSWLKTQOLANRIV--KLNGVLQRT---EQYSYDSNR 1110
DB 477 DHLHREVQR---SQGR-LDMRMYDRTRGLTRKLTCKMRGVVPEFIDREYAYSGQDE 531
QY 1111 LNQYKCDGAECPDXYGHSIVTQNTFYDIYGNITAC-HTTFADGCTEDHATKFNPTDPC 1169
DB 532 LLLK-----KHSRQGVTDYF-YDTTGRTACRNEAYLDSWQYDAA---ANLLDRR 577
QY 1170 Q-----LTEVHHTHPMDPNIRLYKDKAGRVINITDNGHNTENFTYDTL 1213
DB 578 QETAQAGAGSVVPFNRTSYRGLH-----YRYDEVGRVVEKRGNG-TQHYRWDAE 628
QY 1214 GRL-----QNGQGSVYG--YDPLNRLVSQKTDLDCLLYRETMLVNEVRNGEMIRLLR 1265
DB 629 HRLTEVAVTRGTVRRYGYVYDAPGRVKEHLEDAEGKPYNRITFLWDMGLAQECRLGR 688
QY 1266 TGETIIAQOR-----ASKVLLTGTD-SQQSVLTSQKLSQEA-YSAY 1307
DB 689 SSSLYIYSDRGSHEPLARVDRAPQEADEVLYYHTDVNGAPEMTDGGNIVWEAGYQW 748
QY 1308 GK--HKSTANDASILGYNGRADPVSGVTHLNGYRSDYPTLPMRPHPTDLSLSPFG-AGG 1363
DB 749 GNLTHEKETRPVQQLRFQGVLDRETGLHY--NLRYFYDPDIGNFISGD---PIGLAGG 803
QY 1364 INPYSYCLGDRINRSDPSG 1382
DB 804 INLYQYA-PNPLSYIDPLG 821

RESULT 10
US-10-282-122A-47600
; Sequence 47600, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
```

```
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 47600
; LENGTH: 1515
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
US-10-282-122A-47600

Query Match      3.9%; Score 350; DB 12; Length 1515;
Best Local Similarity 22.1%; Pred. No. 4.2e-18;
Matches 232; Conservative 126; Mismatches 314; Indels 380; Gaps 59;

QY 481 YGTRFVKKIQTQYDSEFFDKPEKFIQYRSLIGSQSHVTLKIEERHYSATQLLNSTLF 540
DB 543 FQCFQYVAYDEHGMWMTQRTDQTDVRYRYDTAG-----RVVE-----TGTRO 586
QY 541 QYNTDKSLGRLKQTECKGNGKYSVVKHFTYTKQDDTLQOQSHS-----IITHDNFTI 596
DB 587 CYHT-----GRFIVEAGCT-----RVIDVDGEWTVAYNDEGLVTAETDPLGCHCTYSEWEL 636
QY 597 HRSQVR-----SRV--TGRLPFS-----DTDKOIVT-----QMSY 624
DB 637 GRMARIDPLGRITDYDERGQLTSVVESGRTVDFDYDEQSLTGARLPNGGTIKLEY 696
QY 625 DKLGRLTRTNSGTPVANTLTQY-----EL-----NNQDDNRPPFVITTDV-- 669
DB 697 DHLGRLIART-----EPDCKTYYGRGELLRVVQGDRETRLDYDR-----LRLTDIEL 748
QY 670 -NGNQLRNEFPDGRHVSQCLKSDGDKFYTHIQYDQGRHHTSYSDYLTNGR--- 725
DB 749 PTGARFRKIDALGR-----LLEETSPDG-----HVTRYD-----YADGPANPRGLL 790
QY 726 -CQTDPPDKVHLSMSKSYDNWQCIANTHWSYGVSEKITV---DPITLTATKQLQSNVNVQ 781
DB 791 SAVTRPD-----GSVSRARYN-----SESLPVEWIDPLGRTIQR----- 824
QY 782 TGCEVTTVTPSQPIQITLFDGAGH-----LQSCHTUTRGCWDRVRK 823
DB 825 -----TYGPFDP--LLTASIDAAGHATRFYDHDATRLTKVINALGETYTVRYDAAGSLAA 876
QY 824 ETDAIGQCTIYQYNYNRFVIGITLPDGTIVNRKYPFSTDTLITDIRVNGISLQO--QTF 881
DB 877 EIDWGGRTATYDRDAVGRLLTKLPDG-----GQWRYTY 910
QY 882 DGLSRLTOSQDGRVWAYTYAGNDQCFSTVITPDGQFIHVQYQPELDDAVLQVANSNEIT 941
DB 911 DASDRLEIDAGDVKLAVRYDASG-----RLASAEVQGEHTHTV 949
QY 942 QQFSYNPVYTGALLKAVAGQSILTYIPYSPGLKMNINDMKMSYLVTLRGLNGYTDLT 1001
DB 950 -RFAYDR-NGRLIGDQHGELLRHVYDAGQRRLL-RMTPRETTIYAY----- 993
QY 1002 GTIQXISRDTGRTVQTKDSSIKITPLNYDDLLNRHIGSQVTDL-----ATG-----HMLT 1050
DB 994 -----DVSGALTQVG-----QLTIRSDGLGRIGREAGDFVAQQQYDGLRIRRQIAG 1041
```

```
QY 1051 TTVEPDGLNRBIGRKLCDSCSGHTLIDIOQSWLKTOQLANRIVKLVGLQRTQEQYSYDSNR 1110
DB 1042 PAVAFDALQADPARAL-----BOLTRQV-----YHYDAAGQ 1072
QY 1111 LNOYKCDGACPT---DKYGHISIV-----TONFTYDIYGNITACHTTADGTEDHATF 1160
DB 1073 LERVD-TGADTLTYQDERGQII CAESLQPSHFHYDAVNNI-AAHGQRA----- 1121
QY 1161 KFAPTDPCQLTEVHHTH---PDMPONIRLKYDKAGRVINIT-DNHG---NTENFTYDT 1212
DB 1122 ---PVD-----AHHYRGGGLPEQGVARYKYDARGRTIEKTVQGVPRPKTWQYWDG 1171
QY 1213 LGRL-----ONGQGSVYGYDPLNRLVSQKTTDTCLELYRET-----MLVN---EV 1255
DB 1172 LNRLVVVTPBERGVWA--VRYDAFNRIEKO-----QVGRETVKFLWDGEMLAERWIEQ 1224
QY 1256 RNEMIRLLRTGETI-----IAQORASKV--LLTGTDSQQSVILTSKONLSQE 1302
DB 1225 RDG-----TTGQVVTWIEPESFLPLAQETDDGLFPILTDQIGRPKTVFDEQGRPVKA 1278
QY 1303 AYSAYGK---HKSTANDA-----SILGVNGERADPVSVGTHLNGYRSYDPTLMRFH 1351
DB 1279 AYSLWGLKLPVKRPANDACGATSIDTTLRFSGQWADDETGLNLYNARY---YDPDSGOYL 1336
QY 1352 TPDLSLSPFG-AGGINPYSYCLGDPINRSDPSG 1382
DB 1337 SAD-----PIGLGGARTQAY-VHDFSQWIDPLG 1364

RESULT 11
US-10-282-122A-43059
; Sequence 43059, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/40/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
```

```
; SEQ ID NO 43059
; LENGTH: 1397
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-282-122A-43059

Query Match
  3.9%; Score 342; DB 12; Length 1397;
Best Local Similarity 21.4%; Pred. No. 1.6e-17;
Matches 256; Conservative 157; Mismatches 382; Indels 404; Gaps 63;

QY 418 KTTW-ROADNSYRSEITETTFDESGNPLTKVI-----KDKTKQKIIS--PS----- 461
DB 218 RTQTFHRAAGFSGEITGV--DGAGRHFLVLVTTQAQRAEABEQQAISGGTSPAPDPT 276
QY 462 -----THWEYPPAGEVDNCPPEP---YQFT-----RVK 488
DB 277 LPTTEYGRDNGIRLSAVLTHDPEY-----ENLPAALVRYGWTGPEGELAAYVDSNT 331
QY 489 KLIOTPYSEFK-----DDPEKFIQYRSLIGSOSHVLTKIEERHYSATOLLNSTL 539
DB 332 QVRSFTYDDKYGRMVAAHRTGRPE--ICRYD--SDGRVTEQLNPAGLSVT----- 379
QY 540 FOYNTDKSELGELLKQTEC--TKGNGKTYSVVHKFTYTKQDDTLQOSHSHITHDNFTIH 597
DB 380 YQYEDRTITDLSNRREVLHTQGGG-LKWVKK-----EHADGSVT 421
QY 598 RSQVRSRYTGRLPSTDTKDIWTQMSYDKLGRLLTRTLNSGTPYANTLTYDYELNNLQDD 657
DB 422 QSQPDA--VGRLAQTDAAAGRTTEYSPVVTGLTRIT--TPDGRASAFY----- 468
QY 658 NRPPVITTTDVNGNOLRNEFDGAGRHSQCCKDSD-----GDGKFTYIHTQQVDEQGRH 712
DB 469 NHHSLTATGPDGLERIRREYDEWGRLIQETAPDGDITRYVDNPHSDLPFCATEDATGSR 528
QY 713 HTSYSDYLINGRQOTDPKVLHLSKSYDNMGQIANTHWSYGVSEKITVDPI--TLATK 771
DB 529 KTMWRSY---GQLSFDCSGYTRYDHRFGQVTAHREEGLSQVRAVDSRGLLAVK 585
QY 772 QLOSNNVQTKVETTPSQPIQTLFDEAGHLSCHTLTRDGDWRVRKETDAIGQ-- 830
DB 586 DTQGHETRYE-----YNAAGDLTT--VIAPDG--SRNGTQYDAMGKA 623
QY 831 -CTI-----YOYDYNRVQITLPDGTIVNRKYAPSTDLTIDIRVNGISLGQTF-- 881
DB 524 ICTTQGGITRSMEDVAGRVIRLSENGS-----HTFRYDVLDRLLQIETGFGQRTQRYH 678
QY 882 -DGLSRLTQSDGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEI 940
DB 679 HDLTGKLIRSEDEGLV-----THWYD--EADRLTHRTVNGET 714
QY 941 TQQFSYNPVTGAL--LKAVAGQSLTPIY-YPS-GRUKMENI-----NDM---KMSY 986
DB 715 AERWQYDE-RGWLTDISHISEGHRVTYHYGYDSKRLASEHLTVHHPQTNELLWQHETRH 773
QY 987 LWTLRGLENGYTDLTGTITQIKSRDTHGRVTQIKDSSIKTTLNVDLNRHIGSOVTDLATG 1046
DB 774 AYNAAGLAN--RCIPDSLPAAVEWLYTG-----SGWLSGMKLG 808
QY 1047 HMLTTTVEF--DGLNRIGRKLCDSSGHTLDIQOSWLKTOQLANRIVKLVNGVLTQTEQYS 1104
DB 809 D--TPLVEYTRDLRHRETLR-----SFRYELTAYTPAGQLOSQ--HLNLSLLS-DRDYT 858
QY 1105 YDSNRNLNOYKCDGAECPDKYGHISVTQNTFTYDIGNITACHTTAD-----GTEDHA 1158
DB 859 WNDGELIR-----ISSPRO-----TRSYSGTTRGLTGVTHTTAANLDIRPYTDDPA 906
QY 1159 TFKFANPTDPCQLTVEVHTHPD-----MPDN-----IRLKYDKAGRVINITD----- 1200
DB 907 GNRLLPDP-----ELHPDSALSMPDPNRIARDAHYLVRYDRHGRLTEKTDLIPGCV 956
QY 1201 ---NHGNTENTYTDTLGRQLONGQGSYVG-----YDPLNRLVSKQKTDLDCHEL----- 1244
DB 957 IRIDDERTHYHYDSQRLVHYTRTQYAEPLVSRVLYDPLGRVAKRVRRERDLTGWM 1016

RESULT 12
US-10-282-122A-43060
; Sequence 43060, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43060
; LENGTH: 1411
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-282-122A-43060

Query Match
  3.8%; Score 340.5; DB 12; Length 1411;
Best Local Similarity 20.4%; Pred. No. 2.2e-17;
Matches 287; Conservative 172; Mismatches 459; Indels 490; Gaps 70;

QY 418 KTTW-ROADNSYRSEITETTFDESGNPLTKVI-----KDKTKQKIIS--PS----- 461
```


QY 838 NNRVQITPLPDGTVINRKPSTDTLITIRVNGISLQOQTFDGLSRLTQSDQ-GRV 896
Db 560 RFGOM-----TAVHRE-----EGISL-YRDRNGRLTSVKDAQRE 595
QY 897 WAYTSAGNDQCPSTWITPDQFIHQYQ----- 925
Db 596 TRYENAAAGDL--TAVITPDGNSRSETQYDAWKAVSTTQGLTRSMYDAAGRVISLTNE 653
QY 926 -----PELDAVLQVNASNEITQOFSNPNVTGALLKAVAEQOSLTIYVPSGRLKME 976
Db 654 NGSHSVFSDALDELVLQOQGFDRTORHYD-LTGKLTQSEDEGLVILWYDESDRITR 712
QY 977 NINDMKMSLWTLRGLNGYDITLGTIQKISDTHGRVTQIKDSS-----IKTLLNYDDL 1032
Db 713 TVNGEPAEQY-----DGHGWLTDISHLSEGRHVAHVGYDDK 751
QY 1033 NRHIG--SQVTDLATGML----- 1049
Db 752 GLTGECQTVENPETGELLWQHETKHAYNEQGLANRVTPDLSPPVWLVYGSYLAGMKL 811
QY 1050 --TTTVEF--DGLNREIGRKLCDSSGH--TLDIQQWLKTQQLANR----- 1089
Db 812 GGTPLVYTRDLRHRETVRFSGWAGSNAAYELTSTYTPAGLQSQHLSLVYDRDYGWS 871
QY 1090 ----IVKLVQRTQEQYSYDSNRNLNQYKDCGAE-----PTDKYGHSI----- 1130
Db 872 DNGDLVRAISGPRQ--TREYGYSATGRLESVRTLAPDLDIRIFYATDPAGNRLDPPELHPS 930
QY 1131 -----VTQN-----FTYDIGNITACTTTPADG--TEDHATFKFANPTDPCQLTEV 1174
Db 931 TLVWPNRDAEDAHYVRHDEYGRLTETKTRIPAGVIRIDDET----- 975
QY 1175 HTHPMDPDIRLKYDKAGRVINITD--NHGN--TENFTYDTLGR----- 1215
Db 976 HHYH-----YDSQRLVYTRIQHGEPLVESRYLDPLGRMAKRWRRERDLTG 1025
QY 1216 ----LQOQGSVGYDPLNRLVSOKTDTLQCELYR----- 1247
Db 1026 WMSLSRKEPVWYGMWG-DRUTTVQDTTITQTVYEPGFTPLRVETENGEBEKAQRSS 1084
QY 1248 --ETMLNVRNG-----EMIRLL-RTGETIIA-----QORASKV 1279
Db 1085 LAETLQOEGSENGHGVFPAPLVLRLDLLEEIRADRVSSERAWLAQCGLTVEQLARQV 1144
QY 1280 LLTQDTSQOS-----VILTSKQNLQSEA-YSAYGKHSTANDASILGYN----- 1323
Db 1145 EPEYTPARKAHLYCHDRGLPLALISEDGNTAWSAEYDEWGNQNLNEENPHHV--YQPYRL 1202
QY 1324 -GERADPVSGVTHLNGVRSYDPTLMEFHTPDSLSPPG-AGGINPYSYCLGDPINRSDPS 1381
Db 1203 PQOQHDESGLYY--NRHRYVDPLQGRYITQD---PMLGKGNLYQYPL-NPLQIDPM 1256
QY 1382 GHLNQAWTGMGIMGAGLLLTATGMAAAGGIAAIASTSTTALLAFGALSVTSDITS 1441
Db 1257 GLL--QTWDDARSG-----ACTGV-----CGVLSRIIGSPKFDSTADALD----- 1296
QY 1442 IVSGALEDASPASSIILGWSVMGMAAGLAEBSA-KGGTKLATHLGAFAEDGENALLKSTS 1501
Db 1297 -----ALKETQNR-----LNDMEYSIVCKDING-----KYFASKAETDNLK--K 1336
QY 1502 ESSRIKWGVTSLDREIVRNEBEGQVVKHDSRGYTDNFMWKGEOAILVHGDK--DGLFYHT 1559
Db 1337 ESYP-KRKCPTGTRVAAVYHTHG---ADSHGDYVDEFFSSDKNLVRSKONNLEAFYLAT 1393
QY 1560 EGNKH--NGKGPY 1570
Db 1394 PDGRFEALNNKGEY 1407

RESULT 15

US-09-815-242-10384

; Sequence 10384, Application US/09815242

Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815.242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10384
LENGTH: 1377
TYPE: PR1
ORGANISM: Escherichia coli
US-09-815-242-10384

Query Match 3.7%; Score 331.5; DB 9; Length 1377;
Best Local Similarity 20.8%; Pred. No. 1.1e-16;
Matches 288; Conservative 170; Mismatches 450; Indels 477; Gaps 75;
QY 211 LPPGQKEGY-RTELPLFLNQLNSIHNFSLGNENPLTWSFGYTPICKGILGOWITSMAP 269
Db 126 LPPGE-DGYRSSESMLVR--GGVAKLDEGRLLALWQ----- 160
QY 270 GGLKETVNY-----NNQ-----HFFQSANL--PVLPLYTLMKQVPGAGQ- 310
Db 161 -ALPELRLSPHYRLATNSPGPWLLGCWCVPEADVLPAFLPPYRVLTGLVDRFGT 219
QY 311 PAIQAEYSYTHNYVGGSGNGLWNNKLNLYGLMTEYNYGSTRYKDKGHDQIVRIE 370
Db 220 QTFHREAAAGEFSGEITGYTDGAWRH-----PRLVL-----TQOAR----- 255
QY 371 RTYNNVHLTSCKQNGYIQTETAYYAIIGHNFDSPQSOLFPT---KTETWRSADN 427
Db 256 -----AEARQ-----AISG--GTEPSAF--PDTLPGYTEYGR--DN 287
QY 428 SYRSEITETTFDESNGPLTKVIKOKTKQIISPSTHWEYYPAGEVDNCPPEP---YGT 484
Db 288 GIRLSAVMLTHDPE-----YP-----ENLPAALVRYGWT 317
QY 485 -----REVKKIOTPYDSEFK-----DDPEKTIQYRSLIGSQSHVTLKIE 525
Db 318 PRGELAVVYDRSGKQVRSTYDDKYGRVVAHRTGRPE--IRYRD--SDGRVTEQLN 372
QY 526 ERHYGATQLNLTLPQYNTDKSELGRLLKQTEC--TKGNGKTSYVHKFYTKQDDTLQ 583
Db 373 PAGLSYT-----YQEKDRITITDSDRLREVLFTQCEAG-LKRVVKK----- 413
QY 584 QSHSITTHDNFTHRSQVRSRYTGLFSDTDTKQIVTQMSYDVKLGRLLTRTLNSGTPYAN 643
Db 414 -----EHADGVSQTOSQFPA--VGRLRQATDAAGHTEYSPDVVTGLITRIT---TPGR 462

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2004, 05:44:02 ; Search time 25 Seconds

(without alignments)
6437.137 Million cell updates/sec

Title: US-09-889-874A-23

Perfect score: 8979

Sequence: 1 VYIKFLKLFRRITWSDNNEF.....PRKIILGRTEKTVKPKTRP 1673

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1214	13.7	709	2 S38241	hypothetical prote
2	912.5	10.3	528	2 S38242	hypothetical prote
3	432	4.9	2334	2 S32920	cell wall-associat
4	429.5	4.8	336	2 S38239	hypothetical prote
5	386	4.3	2167	2 AF1489	cell wall-associat
6	381	4.3	2183	2 T37218	hypothetical prote
7	365	4.1	1400	2 E90886	RhsE core protein
8	353.5	4.0	843	2 AB0339	Rhs-family protein
9	342	3.9	1397	2 C64805	RhsC core protein
10	341.5	3.8	1404	2 E90558	RhsG core protein
11	341	3.8	1394	2 H91236	RhsH core protein
12	340.5	3.8	1411	2 E65145	RhsB protein precu
13	339	3.8	1426	2 H64780	RhsD protein precu
14	336.5	3.8	1404	2 E85509	hypothetical prote
15	331.5	3.7	1377	2 C65159	RhsA protein precu
16	331	3.7	1409	2 F91187	RhsA core protein
17	326	3.7	1399	2 A99720	RhsC core protein
18	322.5	3.6	1377	2 E86034	RhsC protein in rh
19	319	3.6	1397	2 A85570	RhsC protein in rh
20	316	3.6	1398	2 B85549	hypothetical prote
21	313	3.5	1398	2 H90598	RhsD core protein
22	297	3.3	1512	2 AH0439	probable membrane
23	280	3.2	1317	2 F83310	conserved hypothet
24	278	3.1	985	2 B86084	hypothetical prote
25	276.5	3.1	1438	2 AI0093	conserved hypothet
26	271.5	3.1	1354	2 AG0538	Rhs-family protein
27	266.5	3.0	2515	2 S47008	tenascin-like prot
28	248.5	2.8	1616	2 E90704	Rhs core protein w
29	245.5	2.8	1645	2 H85554	hypothetical prote

ALIGNMENTS

RESULT 1

S38241

hypothetical protein - Cxiella burnetii

C;Species: Cxiella burnetii

C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Oct-1999

C;Accession: S38241

R;Thiele, D.; Willems, H.; Haas, M.; Krauss, H.

submitted to the EMBL Data Library, October 1993

A;Reference number: S38215

A;Accession: S38241

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-709 <THI>

A;Cross-references: EMBL:X75356; NID:G407370; PIDN:CAAS3129.1; PID:G407397

Query Match

Best Local Similarity 13.7%; Score 1214; DB 2; Length 709;

Matches 287; Conservative 101; Mismatches 280; Indels 36; Gaps 16;

QY	554	KTECTKGNGKTSVVKHTYTKQDDTLQOQSHSTTHDNFTIHRSOVRSRYTGRLFSDT	613
DB	3	RRAEVLTSKEGKKYQNTFEALSOAEHLQKIDFTGEGDKITSISREQSRYSGLLST	62
QY	614	DTKDIVTQMSYDKLGRLLTRILN-SGTPYANLTLYDYELNNLQDNNRPFVITTTDVNGN	672
DB	63	DELGNVTQVEYDELGRLLTQVNASSTTYASTRYSYSLTDDARGKVTAKTTVTDPKGN	122
QY	673	QLRNEFDGAGRHVSQCLKSDSG-----DGKPYTHIQOYDEQGRHHTSTYSYDLTNGRQQ	727
DB	123	QLRTYDGLGRNLKQERLDKDAVSKTGTWTYTHQOQYDALGRESKITITQDVL-----R	177
QY	728	TDPDKVH-----LSMGK--SYDNWGQIANHWSYGVSEKITVDPIITLTATKOLQSNNVQ	781
DB	178	LDSEVGHAGSVLSISKWYHDSWGNHLTVFSDGYQERSVYDPIRRAI--LQESGSQK	235
QY	782	TKGEVTTTPSQPIQITLFDAGHLQSCHTLTRDGMWRVKETDAIGCQTIYQYDNYNR	841
DB	236	LGOQLTEYNLAGLPKVTQYDSQGTQESAHVEYDGLQLRKETDELQIITLYEVDHFR	295
QY	842	VIQITLPGTTVNRKYAPFSDTLITDIRVNGISLGQOTEDGLSRLTQSQDGRWATY	901
DB	296	VTQITLPTNTIIQSYAPHSTASLITGISVNNFMNGNFTDSLRLTETTSGGTSAPSY	355
QY	902	SAGNDQCPSTVITPDGQFIHYQOPELDDAVLOVASNEITQFQSNVPTGALLKAV-AEG	960
DB	356	ENAS-SVPAAVTAPTGETSVSEYKELGNVKKISAPFIIQTWYDALTGMTSATQAAG	414
QY	961	QSLTFIYPSGRKQME-NIND---MKMSVLTWLRGLENGHTDLTGFIQKISRTHGRVT	1016
DB	415	MIQMTYPSGLLNKMTSMFDPGAQAKSTAYTISLAGAPQSYTDFVGTQYDYDEHGRI	474
QY	1017	QIKDSSIKTTILNYDDLNRRHIGSQVTDLATGTHMLTTTTFEFDGLNREIGRKLCDSSGHTLDI	1076

Doc4 protein, stre
probable membrane
hypothetical prote
collagen adhesin -
lactocepin (EC 3.4
adaz protein - frul
hypothetical prote
hypothetical prote
probable peptidog
hypothetical prote
large repetitive p
hemagglutinin/hemo
toxin-like outer m
lactocepin (EC 3.4
IGA-specific metal
cell surface prote

Db 475 GIEDNKLVSLEDAFGRTRKQATDKTKGAVLSLTLYDDLNREIKREISASQSVLVI 534
 QY 1077 QOSWLTKQOANRIVKLVNGVQRTQEQSYSDSRNRNLQYKDCGAECPDTKYGHISVTONFT 1136
 Db 535 ECTYQRNHLKERITQGRGTRTLRKEMFAYDSNRRLIETTCNGEARPDQPFYKAIHQRTFS 594
 QY 1137 YDIYGNITACTTFAAGTETHATFKFANPTDPCQTEVHTHPDMPDNIRLKYDKAGRVI 1196
 Db 595 YDAGLNWTKQIDFSGG-RNTATYIY-SALIDTQLKNNHSDSDVPKXITLEYDKAGMI 652
 QY 1197 NITDNHNTNFYDTLGRQ-NG---QGSVGYDPLNRLVQ 1235
 Db 653 R--DEAGRT--LRDALGRQQVNGAGARGQYAYDVLNLTLSQ 692

RESULT 2
 S38242
 hypothetical protein - Coccidia burnettii
 C:Species: Coccidia burnettii
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 24-Nov-1999
 C:Accession: S38242
 R:Thiele, D.; Willems, H.; Haas, M.; Krauss, H.
 submitted to the EMBL Data Library, October 1993
 A:Reference number: S38215
 A:Accession: S38242
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-528 <THI>
 A:Cross-references: EMBL:X75356; NID:9407370; PIDN:CAAS3130.1; PID:9407398
 C:Superfamily: Coccidia burnettii hypothetical protein

Query Match 10.3%; Score 912.5; DB 2; Length 528;
 Best Local Similarity 40.8%; Pred. No. 6.9e-43;
 Matches 222; Conservative 72; Mismatches 203; Indels 47; Gaps 17;

QY 18 NEF-FTOANFTSAVSGVDPRGTYNIQTILGHIVGNGLGPTLPFLSVPLNKTDI- 75
 Db 2 NELPYQATNFTSAVQGVDPRTGLFTVMVLAELTGNDLGPDLFTFLNYSLSHTSNI 61
 QY 76 GFGIGFNGSVYDRNKSLSSTGENYKVIETDKTKVLOOKLNDLRFKDKLNCYRI 135
 Db 62 GFGIGCSVIGSYDKNKLILISSERYKTEDWDSVYVROKKINPKFEK--IKNGYII 119
 QY 136 IHKSGDIEVLG--FNNAFPLKVPKLLNPAHAIYIDMNFPEATQPLNRIYDDLDGHD 193
 Db 120 KYKNGKTELYNKKYGDNLF--LRQKIFSLGWLPLKTWENRGQYVNLTKIEDAKD--- 173
 QY 194 IPELNLEYOGLIKTILTLFPQKEGVRTELRLNRLNSIHNFSLGNENPLTWFSFGYTI 253
 Db 174 -VLCKIDYQPSDWARITFPFGTESYTPQLDFVNEYLWVTKSTSR--LWMSFNVDV 230
 QY 254 GKNGILGQMITMTAPGGLUKETVYNNYNNQGHHPQSANLPLVPLVTLMQVPGAGQPAI 313
 Db 231 GAGNFT--LTQVKSPTGLTETVNYQAGV--RFPDSGKPALEPVYNYRQSPGQPD 285
 QY 314 QAEYSYTSNHYVGGSN--GIWNNKLDNLG-LMTEYNGSTEGRRYKDKGHDQIVRIE 370
 Db 286 VXEYEYTVSNLYGYGASLGKKNWEDENLYNVMDYYSSTEXKLIVDNR---LVSIS 341
 QY 371 RYNNYHLTSECKQONGIQTETAYVAILGHPDPSQPSQFOLPKYKTETWR--SADNSY 429
 Db 342 RIYNSYLLISITTRQNSCEVITDYAKPGLSPDKQPKQFOLPKKEKTKWRENSKQC 401
 QY 430 RSEITETDESGNPLTKVKKTKQKLIISPTHEWYPPAGEVDN---CPPEPYGTR 485
 Db 402 RSEITTTTDPGNLLTKIEDP-----GKTEYIYDYSKGETDKGVLCPPEPNFVR 454
 QY 486 FVKKIQTQYDSEF-----KDDKEKIQYRSLYSQSHVTLKIBERHYSATQLLNLSLQ 541
 Db 455 FVKTQIVTPADSEFYAPVQQTYYAYAQYPCITAGSLSVAVLQTOETLCSDDVL----LLT 510
 QY 542 YNTD 545

Db 511 INTD 514
 RESULT 3
 S32920
 cell wall-associated protein precursor wapa [similarity] - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 21-Jul-2000
 C:Accession: S32920; E69730; T47101
 R:Foster, S.J.
 Mol. Microbiol. 8, 299-310, 1993
 A:Title: Molecular analysis of three major wall-associated proteins of Bacillus subtilis
 rotein.
 A:Reference number: S32919; MUID:93302506; PMID:8316082
 A:Accession: S32920
 A:Molecule type: DNA
 A:Residues: 1-2334 <FOS>
 A:Cross-references: GB:L05634; NID:9304177; PIDN:AAA22883.1; PID:g304179
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
 C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
 A.; Ehrlich, S.D.; Emmeron, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koester, P.; Koningstein, G.; Kroch, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mausel
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seror
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033; PMID:9384377
 A:Accession: E69730
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-2334 <KUN>
 A:Cross-references: GB:Z99124; GB:AL009126; NID:g2636442; PIDN:CAB15959.1; PID:g2636469
 A:Experimental source: strain 168
 R:Yoshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.
 Microbiology 141, 337-343, 1995
 A:Title: Cloning and sequencing of a 29 kb region of the Bacillus subtilis genome contai
 A:Reference number: Z24350; MUID:95219088; PMID:7704263
 A:Accession: T47101
 A:Status: preliminary; translated from GB/EMBL/DDBU
 A:Molecule type: DNA
 A:Residues: 1-2334 <YOS>
 A:Cross-references: EMBL:D31856; NID:g603765; PIDN:BAA06656.1; PID:g603782
 A:Experimental source: strain BGSC1A
 C:Genetics:
 A:Gene: wapa; N17G
 C:Superfamily: cell wall-associated protein wapa

Query Match 4.9%; Score 432; DB 2; Length 2334;
 Best Local Similarity 20.4%; Pred. No. 4.1e-15;
 Matches 358; Conservative 226; Mismatches 613; Indels 560; Gaps 80;

QY 44 ICITLGHIVG-NGNL-----GPTLPILSVPLNKTDIGFGIGENFGLSVDR 90
 Db 896 IDIFSGQLNGATGNVIVNEEDSLDGRGPGLSRTRNSLSDSHLFGGCV-----YADA 950
 QY 91 KNSLLSLSTGENYKVIETDKTKVLOOKLNDN-----LRFEKDKENYRIIHKSGDIEVL 145
 Db 951 ETSVISTDGGAMY--IDEDATTHRTFKKADGTYQPTGVVLELTETADQFILTKD---- 1004
 QY 146 TGFNNNAFDLKVPKLL-----NPAGHAIYIDWN-----FEATQPLNRIYDDLDGH-- 192
 Db 1005 ---QTNAYFNKGGKLOKVDGHNATVYNDKNOITAITASGRKLTFTYDE-NGHVT 1060
 QY 193 -----DIPLNLEY--QGLI-----KYTLTLPFG 214

Db 1061 SITGPKKKVYTSYENDLLKKVTDITDGTSTVDYDSEGLRVKQYSA NSTAKPVFTY-- 1118
QY 215 QKEGYRTELRLNQLNHSINFS-----LGNENPLTWGFTYPIGKNGILGOWITSMT 267
Db 1119 QYSGHLEKALNAKKEVTVYSYDADKKTLLMQPNGRKVQYGYNEAGNP-----IQVID 1172
QY 268 APGGIKETVN--YNNN-----OGHFFPQSANLPVLPIYTLMMKQVDPGAG 309
Db 1173 DAEGELKITNTYKGNVVEVDNDVGTGKATESYQYDKGN-----VTSVKDAYGT- 1225
QY 310 QPAIQAEYSYTHNYVGGSGNGIWNKLDNLYGLMTEYNGSTESRRYKDKEGHDQIVRI 369
Db 1226 -----ETYEYNNKNDV-----TKMKDTEGNTVDIAVDGLDAVSETQSGKSSAAV 1271
QY 370 ERTYNNHLLTSECKQONGYIQTETAYAILGHNFPSQSPQFOLPKTKETETWRSADNSY 429
Db 1272 YDKYGNQIQSSKOLASASTNLK-----DGSFPAQSGMNLTASKD----- 1311
QY 430 RSEITETTFDSGNPLTKVINDKTKTKIISPS-----THWEYPPAGVDNCPPEPYGFT 485
Db 1312 RKKIS-VIADSG-----VLSSKALEVLQSSTAGTDHGYSSATQVLEPLENTYLSG 1365
QY 486 FVKKII---QTPYSEFKDPEKFIQY---RYSLIGSQSHVTLKIBERHYSATQLLNS-- 537
Db 1366 KIKTDLAKRAYFENIDLRDOKRIOWIHNEYSALAGKNDWT-----KRQITFTPANAGK 1421
QY 538 ---TLFOYNTDKSELGR-LKQTECTKGNGKTVSVVHETKYTKQDDTLQOS-HSITTD 592
Db 1422 AVYMEVDHDKDQKGAWEDEVQLEGEVSSINPNQNSFTSATENMVNSGASVDSEE 1481
QY 593 NFTTHRSQVRSRYTRGLFSPDTDKDIVTQMSYDKLGLLRT-----L 635
Db 1482 GFNDVSLKAARTSASQAGSVTKQTVVLGQSANDKPVVLTITGMSKASSVKFTDKDYSL 1541
QY 636 NSGTPYANTLYDYEL---NNLQDNRPPFVI----- 664
Db 1542 QANVTYAGSGTGIYNAKFPSTGDEWNRAAVVIPTKPKINKVD:SIILFKSATSATVWFPDI 1601
QY 665 -----TTTVDVNGNQLNFPDAGRHVSQCLKSDGDKGFYTIHTQYDEQGRHHT 714
Db 1602 RLIEGSLTKSTYDSNGVYVKEDELGYATS---TDYDTGK-----KTSBTDAGEKTT 1654
QY 715 STY---SYLNGRQOTDPDKHLSMSKSYDNWG-QTANT-----HWSYGVSKSI-- 760
Db 1655 YTDQADQLNTMLNNGTSILH-----SYDKEGNEVSKTRAGADQTYKFEYDVMGKLVK 1709
QY 761 TVDPITLTATKQLQSNNVOT---GKEVTTTPSQPIQITLFEAGHLQSCHLTLD 816
Db 1710 TTDPLGNVLASEVDANSNLTKTISPNGEV-----SLSYD 1744
QY 817 GWRVRKETAIGCTTIQYDYNRNVIQITLPGTIVNRKYAPFSTDLITDIRVNGISL 876
Db 1745 GTDRVKSXNGTEKYIFTYDKNGN-----ETSVVNKEQN-----TT 1781
QY 877 GQOTDGLSRLTQSQDGRVWATYSAGNDQCSTWITPDQFIHYQYQPELDDAVLQVA 936
Db 1782 KKTFNKNRLTELTRGGSQWTYPSDSKLKTF-----SWIH-----G 1821
QY 937 SNEITQFSYNPVTGALLKAVAREQSILTPYPSGRLKMNENDMKMSYLMWTLRGLNG 996
Db 1822 DQKGTNQFTN-----KLDQIEMKDSISSYFYDEN- 1854
QY 997 YDLTGTI QKISDTHGRVTOIKDSIKITLNVDDLNHRHIGSQVTLATGHMLTTTVEPD 1056
Db 1855 -----GNVQ-----TFTGNGGGTFSFYDERNLVSSLHIGDKNGDILTESYEY- 1898
QY 1057 GLNREIGRKLCDSSGHTLDIQQSWLKTQOLANRIVKLVNGVLQRTQYSDYDSNRNLNOYKC 1116
Db 1899 -----DANGRTTINS-----ASKV-----OYEYKLNQLVK----- 1927
QY 1117 DGAECTDKYGHIVTONFTYDIYIGNITACHTTFADGTED--HATFKFANPTDPCQLTEV 1174
Db 1928 -----ETHEDGTVI---EYTDGFGNRKTV-TTIKDGSSKTVNASPNMN-----QLTKV 1973

QY 1175 H-----HTHP-DWPDNI-----RLKYDKAGRINUITDNHG 1203
Db 1974 NDEISYDKNGNRTSDGKETTYTWDADENLTATVKKGEKFPATYKYDEKGNRIQKTVN-G 2032
QY 1204 NTENFTYDTLRLQNGQGVYDPLNRLVSOKTTDTLDCELY-YRETMLVNEVRNEMIR 1262
Db 2033 KVTNYFYDG-----DSLNLVYETDADNNVTKSYTYGD-----SQLLS 2070
QY 1263 LLRTGTETIAQORASKVLLTGTDSQSVILTSQKQLNSQAEYSAYGK-HKSTANDA---S 1318
Db 2071 YTEGKKYFYHYNAHCDIIAISDSSTGKTV-----AKYQDANGNPTKTEASDEVKDN 2122
QY 1319 ILGYNGERADPVSGVTHLNGYRSYDPTLMRPH--PDSLSFPFGAGGINPYCYCLGDPIN 1376
Db 2123 RYRYAGYQYDEETGLYLMARY--YEPRNGVFLSLDPPGSDGSDLDQNGYAYGNPNPM 2180
QY 1377 RSDPSGHLSSQAWTGIWGIGIAGLLLTATGGMAIAAAG---GIAAAATASITTTALAGA 1432
Db 2181 NVDPDGH--W-VLWLVNAGFA-----AYDGYKAYKSGKWKGAWAASNFGPKIFRG 2231
QY 1433 LSVTSDITSIVSGALEDASPKASSILGWVSMGV---GAAGLAESAIRKGGTKLATHLGAF 1488
Db 2232 ASRAYKFTK-----KAVKITGTHRHGLNQSIGRNG-----GRGVNLRAKLNA- 2273
QY 1489 AEDGENALLKSTSESRIKGVYTRSLDRB--IVRNEGOVVKDHSRGYTDNPMGKGBOAI 1546
Db 2274 -----VRSPKVKIQPNGATKYVGKATVVLNKRKGVITAYG-----SS 2312
QY 1547 LVHGDQDGFVHYTEGKN 1563
Db 2313 RAKGSKHVPHTHGKGNK 2329

RESULT 4

S38239
hypothetical protein - Coccidiella burnetii
C:Species: Coccidiella burnetii
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Oct-1999
C:Accession: S38239
R:Thiele, D.; Willems, H.; Haas, M.; Krauss, H.
submitted to the EMBL Data Library, October 1993
A:Reference number: S38215
A:Accession: S38239
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-336 <THI>
A:Cross-references: EMBL:X75356; NID:g407370; PIDN:CAA53127.1; PID:g407395

Query Match 4.8%; Score 429.5; DB 2; Length 336;
Best Local Similarity 34.3%; Pred. No. 2e-16;
Matches 121; Conservative 48; Mismatches 91; Indels 93; Gaps 13;
QY 1348 MRFHTPDSLPFGAGGINPYCYCLGDPINRSDPSGHLSSQAWTIGMGIAGLLTIATGG 1407
Db 1 MRINCPDSWSPFGAGGINPYACDGPINRVDPNPHLSQWAEIDIGLVGLVLAFTAG 60
QY 1408 MAIAAGGIAAAIAASTTATAFALSVTSDITSIVSGALEDASPKASSILGWVSMGMA 1467
Db 61 TSIAAGAGISAAIESASISLVVGTGLVAADVASIASGALEDANPQASATLGLISLUGG 120
QY 1468 AGLAE---SAIKGGTKLATHIGAFABDG-----ENALLKSTSESSRIKMGWT--RS 1513
Db 121 PGAVSGLATAARAGKKL---ISGLAKGGKIRSQSPVQGI-SYRSLRSGDPLRGPPHPQS 177
QY 1514 LDREIVRNEE-----GOVVKDHSRGYTDNF-----MGK 1541
Db 178 LSRVTVAPEMRPAGLNWYHKVSKSSLGVOHVFCAADREIFGYEIRPIEFRRRRPITK 237
QY 1542 GEQAIL-----VHGDKDGFLYHTEG-----NKHNGKGPYTHHTPEQLVDY 1581
Db 238 RDI VILSGTHGRVHGDN---WTSOGLRRRPDILERAFFIEDVQNYKG-----QLNGR 285

Qy 1582 LKONNIVDLTQGG-----DKPVHLLSCYKSSGAADKMAKYIN-RPVIAV 1625
Db 286 KVIVDMAGTSEFGRHVRNVNQHVLGYCWRGDEA---LTYHRNLRPVISY 335
RESULT 5
AF1489
cell wall-associated protein precursor wapA (B. subtilis) homolog lin0454 [imported] - I
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_charge 27-Nov-2001
C:Accession: AF1489
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitouran, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AF1489
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2167 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC95686.1; PID:G16412895; GSPDB:GN00178
A:Experimental source: strain clipl1262
C:Genetics:
A:Gene: lin0454
Query Match 4.3%; Score 386; DB 2; Length 2167;
Best Local Similarity 19.4%; Pred. No. 1.3e-12;
Matches 327; Conservativity 212; Mismatches 582; Indels 568; Gaps 67;
Qy 17 NNEFTQANNTSAVGGVDPRTGLY-----NIQITLGH-VGNGL-----GP 59
Db 718 NNTTSVQEPATPSIPEALNKQGVDPVTSVPVRGGEVNATNGFHTDTNLEGRGP 777
Qy 60 TLPLTSLVPLNKTDIGFIGNFGLSVDRKNSLLSLSTGENYKVIETDKV----- 112
Db 778 SINVRTNSQDDATGIFGKWTSTL-----EEKLVEENG-NIVWVESDKIHRFTKKG 831
Qy 113 -----KLOQKLD-----NLREKDKLNCYRIIHKSGDIEV 144
Db 832 DRYEAPPPIVSEITKNADGYLKBEDKSETFLVDGLKSEKOTKGNELTYEYDGLKTS 891
Qy 145 LTGFNNAFDL-----KVPKLLNAGHAIYIDMNFATQPLNRIYDLDGHDPLNLE 200
Db 892 LRDAAGRTVTLTYEGELKVELGVEDRKLISYTN-----D 926
Qy 201 YQGLIKTILTLFPQOKEGYRTELAFNLNQLNSIHNFSIGNENPLTWSFGYTPIGKNGILG 260
Db 927 KQELISSSTARGKLYRYGTDGL-----LTSYDPKHKTEEKPYETTFAYEE----- 972
Qy 261 QWITSMTAPGGLKFTVYNSNNQ-----HHFPOSANLPVLPVYVTLMKQVPGA 308
Db 973 EKLTSITDPVGKKTLSYDKABQOTTLTNEKKKTIYSNDAGN-----PKKEIVDA 1024
Qy 309 COPAIOABYSYTHSYVGGSGNGIWNKLDNLYGLMTEYNYGVSTESRYKDGEGH----- 363
Db 1025 DGLKLTITYYESNN-----LVKEVNPKGQEEYVAYDADGNITKAT 1065
Qy 364 DOIVRIETNYNHYLLTSECKQONGYIOTTETAYVAIIGHNFDSPQFQLPKTKETWR 423
Db 1066 DAYGESYINDNDVTS-----STDE-----GRKTTYD 1097
Qy 424 SADNSVRSBITP-----TTFDESGNPL-----TKVIK 450
Db 1098 GADAVSETLATESQVSSVTOYDAYGNPIRGSELSSGGLNQLNSGFEKAGVSNWTLIQS 1157
Qy 451 DKK-----TKISP-----STHWEYYPAGEVNDNCPPEYGTFRVKK 489
Db 1158 DAKGSMTFDNTQAPCALGGSGVKLTSEANSVTKGYSVTRQVDVEPETTYTSMIKT 1217
Qy 490 -----IIQTPYDSEFKDDPEKFIQVYSLIGSQSHVTLKIEER-HYSATQLLNSTL 539

Db 1218 SGMNADALLIGRLQDANAKDVTDAVW-----QSNRATSIKNGDWVQRQL----- 1264
Qy 540 FQYNTDKSELGRLL-----KQTECTKGENGKTYSVVHKFYTKQDDT 581
Db 1265 -TFXTSKNTRQVLLYLDNEQAPHKGXTIWDNVQPEKGSVASSYVNVNPFENHNGT 1323
Qy 582 LQOSSHITHDNFT---IHSOVRSRYTGRFLSDTDTKDIVTOMSYD-----KLGRLLTR 633
Db 1324 LPTGMRITGNLTALQAKVVDNQHSHGDSANVFERKATSEAYTHIVQDVPVQKAEKALTI 1383
Qy 634 TLNSGTPYAN-----TLTYDYEL-----NNLQ-----DDNRPPFVIITTD 668
Db 1384 SALSKEDEKANGSVATMSNDYSVMGTYYVQDGTTSVQGFPLGTITDMNRSVAVVWPTK 1443
Qy 669 -----VNGQL-RNEPDGAGRHVSQCLKSDGDKGFYTI 701
Db 1444 PVKIKVYTMFRNGLTGKAFDDVRVIEGEVLTTKNEYDASGVY----- 1497
Qy 702 HTQOYDEQGRHHTSYSDYLTNGRQOTDPDKVHLSMSKSYDNMGQOIANHWSYGVSKIT 761
Db 1498 -TASYDEGRKISFTYDIY---GNKSETDEKGNKKTLYDADNALIDTTLKANGTSVAYK 1543
Qy 762 VDPITLTATKLOQSNNSNVQKVEYTFPSQOI-----QITLDEAGHLQSCHTLT 814
Db 1544 YD-----DNGNTTEKNVTASGKTOKNIYEVVDNKKITAFIDA-----LNRTI 1585
Qy 815 RDGMDRVRKETDAI---GOCTIYQVDNVRNVIQITLPGTIVNRKYAPFSTDLITDIRV 871
Db 1586 KEYDAAGNEFKALIMPGRVTESTYDSADRMKGKWDKLAFFQYDPNGNQTVD-EI 1644
Qy 872 NGISLQOQTFGLSLRLOSQDGRVWAYTYSAGNDQCPSTVITPDGFIHYQYQPELD-- 929
Db 1645 NSI-VTDKTYDDANRITKVAERGGDSVT-----YDKPETKDNK 1682
Qy 930 ---DAVLQVASNE---ITQOFSYNPVTGALLKVAEGOSLTPYVPSGRUKMENIDMK 982
Db 1683 GKTCKVGEVAINHGDTYAKTSYTYNDL-----RNRVNDGS 1719
Qy 983 KMSYL-WTLRGLNGYTDLTGTIQKISRTHGRVTOIKDSSIKTTLNVD-LNRHIGSQV 1040
Db 1720 XNAYFEFDFGNGINVTAGNTAANYTDS*QKVTAANAISASGTQILDENYTYDAASN 1779
Qy 1041 TDLATGMLTTTVEFDGLNREIGRKLCDSSGHTLIDQOSWLKTOQLANRIYKLVGLQRT 1100
Db 1780 TSIDNKKQDGKTYEYDAVNO-----LTKETLPDGTVK----- 1811
Qy 1101 EOYSYDSNRNLNQYKCDGAECPT--DKY--GHSIVTQN--FTYDIYGNITACHTTFADG 1153
Db 1812 -AYTYDGNRTQVAISGSETKTIDASYNQGNOLVSNWGEALTYDANGNRT-----SDG 1864
Qy 1154 TEDHATFPANPTDPCQLTEVHTHPDNPDIRLXYDKAGRVINTDNGHTENFTYDIL 1213
Db 1865 KTYT-----WDTGDRUSSIT-KKGSEPT----- 1889
Qy 1214 GRNQGGSVGYDPLNRLVSKQTDITLDCELY---RETMVNEVRNGEMIR--LLRTGE 1268
Db 1890 -----SYTYDDNRLSKTVGDTTNYHYDGDSDIVLYETDGDGKVRQYVYSDDN 1940
Qy 1269 TIAQORASKLLTGTDSQOVSILTSKQN--LSEAYSAICK-----HKSTANDAS 1318
Db 1941 VRLAMQNGKTLTYHYNAGDVIALTDEAGKIVAYAYDAGNVLKNTASTAEKANKPYG 2000
Qy 1319 ILGYNGERA-----DPVSGVTHLNGYRSDPTLMFHTPTDLSLSPGAGGINPY 1367
Db 2001 YAGTYDKIEIQYLMARYVEPEQV-----FTAYDP-----YFGDEDDP---QTMWY 2046
Qy 1368 SYCLGDPINRSDPSGHLW---QAWTGNGIAGLLIITATCGMAIAAAGGIAAIASTS 1424
Db 2047 NYANNPNVMMFPDGNVAVWIAAAGYGAPEGAEYLLTTKKRWK-----GFGKAVVKA 2101
Qy 1425 TITALAFGAL 1433

Db 2102 VLGLGFGKL 2110

RESULT 6

T37218

hypothetical protein SC2H4.02 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T37218

R:Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, September 1998

A:Reference number: Z21615

A:Accession: T37218

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2183 <OLI>

A:Cross-references: EMBL:AL031514; PIDN:CAA20596.1; GSPDB:GN00070; SCOEDB:SC2H4.02

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC2H4.02

Query Match 4.3%; Score 381; DB 2; Length 2183;

Best Local Similarity 19.8%; Fred. No. 2.5e-12;

Matches 366; Conservative 221; Mismatches 669; Indels 594; Gaps 91;

QY 99 TGENYKVIETDKTKVLOOKK-----LDNLFKEDLKENCYRIIHKSGDIEVLTFP 148

Db 620 TGDG-----TQSLWLDIKHTKRGTDLSLDVAKFSHVMPN--RVGFSDDI----- 666

QY 149 NNNAFDLKPKKLLNPAGHAIYIDW-----NFEATQRLNRIYDDLDGHDPI 196

Db 667 --LSFERPLRTVTVSETGAQITVDYLPADCVAGQTMKADENTKRCPPVYWSVYQEEPI 724

QY 197 LNLEYQGLIKTILTFP--GOKEGYRTELFLNRLNLSIHNSLG-----NENPL----- 244

Db 725 LDWFQKYPVSVRTDPLGSEA-----VQHYEYSGGAWHYNDDELTPAKER 773

QY 245 TWSGYTPIGKNGILGW-----TSMTAPG-----LKETVYNSNNQGHHPQSANLPVL 296

Db 774 TWS-----QWRGYKVTHYTPGSGGTRAKTVTVYLRGMDGDRVLGDDGDKTPD 820

QY 297 PYVTLMKQVPGAGAPAI-----QAEVSYTHNYVGGSGNGLNWKLDNLYGLMTE 346

Db 821 PDRRKAEVSGVTAATDSQLAGFQREVSVYDGDKEVSGTVNDPMSKRT----- 871

QY 347 YNKGSTESRRYKKEGHDQIVRIERTYNNVHLLTSECKQNGY--IQTETAY--YAIIG 402

Db 872 ----ATQKHSVADTEAY--YVRGAS-----HARTRITSRLNPYDRVTRTIGYDDYGMFV 921

QY 403 HNFDSQPSQFLPKTKTETWSADNS-----YRSEITETFD-----ESGNPL 445

Db 922 SVBGLGDDSVTGDEKCTRITWYARNDAAGLTALVSRTIYGRAGSVTDSALDLPADAARP- 980

QY 446 TKYIKDKKTKQKIITSPSTHWE--YYPAGEVDNCPPEPYGTRFVKKIIOTPYDSEFKDDP 503

Db 981 GDVVSOKAT---AYDITNSATQKPYKGD--RWSGRAKYGDDRDPLQWQTAVDY--DTL 1035

QY 504 EKIQYRYSILIGSQSHVTLKIEBRYSATQLNSTLFOYNTDKSELGRLLKOTECTKGEN 563

Db 1036 GRPQVQR-----NTDDVITSKTEYQPPAAGPLTSTTVV-----N 1069

QY 564 GKTVSVVHKFTYTKQDDTLQOQSHSITHONFTIHRQSVRSRYTGRLFSDTDKDIYQMS 623

Db 1070 AK-----GHRTTVKDFAL-----GADLKVTDANGKRVTESA 1100

QY 624 YDKLGRLLTRTL--NSGTPPYANTLYDYELNNLODDNRPFPVITTDVNGNQLRNEFDGAG 682

Db 1101 YDSLGRVTSWLEPNRGRALKGTANYVYGV--VKSTSLFWSSAT--LNG-----DSGS 1150

QY 683 RHVSQCLKSDGDKGKYTHITQYDEQGRHHSTISYDYLNGR----- 725

Db 1151 YRTTYEYDS-----LLRTRQVQAPSAQGGKRVIAQTLYDGRGLFVTAQADINDDTAAP 1203

QY 726 ----QQTDPDPKVLHLSMSKSYDNWGOIAN--THWSYGVSE----- 758

Db 1204 AGKIVQIDGGCAPQRTDSVYDGMGRVTYKAVTKSYGVTEWAVDTAYRGDLVLTCAPEGGSA 1263

QY 759 -KTVDPITLITATKQLOSNNVOTKEVTTYTPSOOPIQITLDFEAGHLQSCHTLTRDG 817

Db 1264 NAVVDAFGRTVERRDYAGTQAPAGTDMYTRYA-----FDAADRKSTAHDRSA 1313

QY 818 W-----DRVEKETDAIGQCTIYQYDNNVNRVIQITLTPD 849

Db 1314 WTYTYDLFGHVSVPDPKGTITVTEVDALDRAVKSTDGREVLFFYDLVGRKTGMWQSA 1373

QY 850 GTIVNRKYAPFSTDTLI-----TDIRVNGISLQ-----QTPDGLSRLTQSC----- 891

Db 1374 KTCAN--KLAASPDITLAKGQODTAVRYEGGETGRAVTKQVTRYDPLYKVYNNELTLPAND 1432

QY 892 ----DGRVWAYTVSAG-----NDQCPSTVITPDGQFHYQYQPELDDAVLOVASNEIT 941

Db 1433 PLVAAAGVPARLAFSTGYNLGDTGVKQAAAPAVAGLSAETVSYTY-----DGLQVLTAKGT 1487

QY 942 ----QOFSYNPVTGALLKAVAEQSLTPIYFSGRLKMNINDMKMSYLVTLRGLNGY 997

Db 1488 TGYLQQAAYSP-----LGLRQWTLATDPTGAKKVLNND-----YEAGTRRLTRSY 1534

QY 998 T--DLTGTI---QKISRDTGRTVQIKDSSIKTLLNYDDLNHRHIGSQVTDLATGMLTAT 1052

Db 1535 VTDDVEGFMLQELKYQQDDAGNITSVSDA---TTLG-----GTGRADHQC 1576

QY 1053 VFEDGNLRNRIGRKL-----CDSSGHTLDIQOSWLKTOQLANRIVKLNGVLQRTQSYD 1106

Db 1577 FTYDG--HRELSEAWPETADCSSTGRT-----VAGLGAAPYWTYSYQVD 1619

QY 1107 S---RNRLNQYKCDGAECPD--KYG-----HSI---VT---QNTYDIYGNITACHTT 1149

Db 1620 DGLRSKQTEHRMSGDDVTTEYEGTABQOPHALSATVTGAENASYTYDETGN----- 1672

QY 1150 FADGTEDHATEFANPTDPCQLTEVHHTHPDMDNIRLKYDKAGVINITDNHGTENFT 1209

Db 1673 -----TE---TRPGVRAQTLDWNAEGRLAGVSEPAAGCKPAT 1707

QY 1210 YDTLGRLONGQSVYGYDPLNRLVSKQTDTLDC-----LYYRETMVYNEVR--NGEM-----I 1261

Db 1708 -----GTAYVYDAGGDLIRRPPTTDDGETVLYLGTTEVHLKVSNGAALKSGA 1756

QY 1262 RLRTGETIIAQRASKVLLTCTDSQSVILTSKQNLQSEAYSA-----YKG 1309

Db 1757 RYTKAGSAVIAVRTST---AGVSGTKLTFLAGDHHGTSGLAINDATLAFKRWSPTFGA 1812

QY 1310 HKSTANDA--SILGYNGERADPVSGVTHLNGYRSYDPTLMRFHTPDS--LSPFGAGGINP 1366

Db 1813 PRGTASGAWPDDRGFLGKPADAAATGLTQ---GARQYDPTDTCGRFLSVDPLEPKENTLNG 1870

QY 1367 YSYCLGDPINRSDPSGHLSSQAWTIGMGIAGLLLTIAT--GGMAIAAGGIAAIAST 1424

Db 1871 YAYASNSPVTNSDPG-----TSDGLG--GLLGAIGAILGGVVGVAVGAVGAIA--TA 1919

QY 1425 TTAALFAGALSVTSDITISVGALEADSPKASSILGWSMGMGAAGLASSALKGKTGLATH 1484

Db 1920 VSLGGGG-----GGWGGTAPTSSG--GW-----TOPLTKQWTPGATY 1956

QY 1485 LGAFADENALLKS-----TSES-----SRKMGVTRSLDRREIYRNEE----- 1523

Db 1957 -----NFTKSWDLFPNPPSQSLEEMLASMPDWGIVS--DPKAANRWETSRSLPFG 2005

QY 1524 ----GOVIKDS--RGYTDNFMKGQQAIIIVHGDK--DGFLYHTEGKNKNG--KGPYTRHT 1574

Db 2006 WLWGGGFLRHEQDFRG-----GDATSIQAQDETISGLRSKMVGQARKGTGKFAKEV 2059

QY 1575 PEQLVYDLKNNIYDLTQGGDKPVHLLSCYKSSGAADKMAKYNRPVI----- 1623

Db 2060 GFQYVD-----EGPEPGSPWKFNLSRGAAKDIAGVLNAGVGTGNOADAF 2107

QY 1624 AYSNKPTISQGLASIERKDFLKSTYSYDPRKIILGRTEKTVKPKTRFP 1673

Db 2108 TYSKARIS---INKKESVTLKPSAWGSD---WSATHVPRSWNP 2150

RESULT 7

E9086

Rhs core protein with extension [imported] - Escherichia coli (strain O157:H7, substrain C); Species: Escherichia coli

C; Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001

C; Accession: E9086

R; Havaashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A; Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic islands

A; Reference number: A99629; MUID:21156231; PMID:11258796

A; Status: preliminary

A; Molecule type: DNA

A; Residues: 1-1400 <HAV>

A; Cross-references: GB:BA000007; PIDN:BA035484.1; PID:gl3361527; GSPDB:GN00154

A; Experimental source: strain O157:H7, substrain RMD 0509952

C; Genetics:

A; Gene: Ece2061

C; Superfamily: rhsf protein

Query Match 4.1%; Score 365; DB 2; Length 1400;

Best Local Similarity 20.9%; Pred. No. 8.9e-12;

Matches 284; Conservative 154; Mismatches 467; Indels 456; Gaps 65;

QY 418 KETW-RSADNSVRSHEITTTDESGNPLTKVI-----KDKTKIISFTWEY 466

Db 218 RLTYREAAAGLAGEITGV-DGAGREFRLVLTQAQAEARKQHTASLSPDT--- 272

QY 467 YPAGEVDNCPPEPYGFTFVKIKIQTVPDSEFKDDPEKFTQYRSLIGSOSHVLKIE 526

Db 273 -----RPLSDSAFPDPLPGTEY-----GPDGIR----- 297

QY 527 RHYSATQLNLSLTFQYNTDKSELGRLLKQTECTKENGKTVSV-----VHKETYTKOD 579

Db 298 ---LSAVLTHDPAYPELPGALARY-----TYTEAGELLAVDRNTQVRAFTYDAQH 349

QY 580 DTLCQSHSITTHDNFTIHSQVRSRYTGRFLFSDTDTKDITVQMSYDKLGRLLTSLNSGT 639

Db 350 PGRVAH-----RYAGR-----PEMYRYDDTGRVVEQLNPAGL 383

QY 640 PYANTLTYDYELNLODDNRPFFVITTTD-VNGNQLRNEFDGAGRHVSQCLKSDGDKGF 698

Db 384 SY-----RYQYE-----ITVTDLSNRREVLTGGAG--LKRVVYKELADG-- 425

QY 699 YTIHQYDEQGRHHTSTYSYDLTNGR--QOTDPPDKVHLSMKSQYDNWGWQIANTHWSYGV 756

Db 426 -----SVTHSGYDAAGRLTAQTD-----AAGRTEYGL 453

QY 757 SEKITVDPTLTATVQLQSNNSWOTKEVITYPSQOPIQITLFDAGHLQS-----CHT 812

Db 454 N--VWSGDITDITPDGRETKEFYNDGNQLTAVVSPDGLSEKRAYDPFGLVSTSCGD 511

QY 813 LTRDGDVRKE-----TDAIQCCITQYDNNRVIQITLDPDGVINARKYAPFSTDLIT 867

Db 512 VIRAYDNPHELPAITTDATGSTQMTWSRYGQLLAFTDCSGYQTRYEYDFRQMTAVH 571

QY 868 DIRVNGISLQOFTDGLSLTOSQG-GRVWAYTYSAGNDCPSTVLTDPQPHYQYQP 926

Db 572 ---REEGIS-RYARYDNRGSLTSVKDAQGHETRYEYNAAGDL--TAVITPDGNSRETOY-- 624

QY 927 ELDDA---VLQVASNEITQQFSYNPVTGALLKVAEGOSLTPPIYPSGRKLMMENDMKK 983

Db 625 ---DAWGKAVSTQGGSLTRSEYD-LAGRIITLTNENSGRSEFTYDA----- 667

QY 984 MSYLTWLRLENGYDITCTQIKISRDTHGRVTKDKSSIKTTLNYPDLNRHIG----- 1037

Db 668 LDRLVQQRGFD-----GRTQRYHYDTGKLTQSGDEGLTLLWHYDESRLTHRTVNGE 720

QY 1038 -----SOVTLATGHMLTTTVEFDGLNRBGRKLCDCSSGHTLDI-----QQS 1079

Db 721 PAEQWQYDEHGWLTJESHLSGHOVAVHYGDDKRLAGERQTVHNPETGELLQWHEH 780

QY 1080 WLKTOQLANRIVKLVLTQTEQYSYDS-----RNRL-----N 1112

Db 781 AVNEQGLANRVTTP--DSLPRVWLTIGSYGLAGMKLGGTFLVETFRDLRHRETVRFGNN 838

QY 1113 QYKCDGAECPDCKYGH-----SIV-----TONFTYDIYGNIT 1144

Db 839 AVELTSTVTPA---GHLSQORLNSOVYDRDYDNDNDGLVRISGPRQWEGYSATGRLE 895

QY 1145 ACHTTFADGTEHDATFRFANPTDPC--OLTEVHHHTPD-----MPONIRL-----KY 1189

Db 896 SVRTLASD-----LDRIPIYATDPAGNRLPD--PELHPDSTLTAWPDN--RIADAHVYVRH 948

QY 1190 DXAGRVINITD-----NHGNTENFTYDTLGR-----NQGQSV---YCYDPL-- 1229

Db 949 DEYGLTEKTRIPAGVIRTDDETHHYHYSQHLVFTYTRIQHGEPLVESRYLYDPLGR 1008

QY 1230 -----NRLVSKQTDLTDLCELYR----- 1247

Db 1009 RMAKVRWRERDLTCWMSLSRKPEVTWYGDGDLRTVQDTTRIQTVYEPGFTPLIRV 1068

QY 1248 -----ETMLYNEVRNG-----EMIRLLRTGETIIAQORAS----- 1277

Db 1069 ETENGEREKAQRSLAETLQOEGSENGHVVPFAELVRLDLRLEERADRVSESRAWL 1128

QY 1278 -----KVLTTGTDSQ--QSVLTSDKQNLQSEAYSAVGHKST 1313

Db 1129 AQCGLTVEQLARQVPEYTPARKVHFYCHDRGLPLALISEDNTAWRGYDEWGNQNE 1188

QY 1314 ANDASI---LGNGERADPVSGVTHLGNYSYDPTLMRFHTPDSLSPFG--AGGINPYSY 1369

Db 1189 ENPYVLPQYPLPQOQHDEESGLYNNRY--YDPLQGRYITQD---PIGLAGMWLYNY 1243

QY 1370 CLGPIINESDFSG-----HLSQAWTIGMGIGAGLLT-----IATG---GMAIAAGGI 1416

Db 1244 PL--NFIIRMDPLGLNLYQLLYDWHDDSYGTSIDITSGDILISLGHAGLVAFKKK 1302

QY 1417 AAATIASTTTALAFGALSVTSDITSIVSGALEDASPKA--SSILGWVSMGMGAAGLAESA 1474

Db 1303 GEMLSIDICIYATACCHAGIGGINAAITYSYKSLPTSGVSNVSG-VTVGGVGG----- 1356

QY 1475 IKGCTKLATHLG-AFAEDGENALLKSTSESRKIKWGVTRSL 1514

Db 1357 -----HPAYIVVDVNDP--ESSTESVGGICAGVDASV 1386

RESULT 8

AB0539

Rhs-family protein [imported] - Salmonella enterica subsp. enterica serovar Typhi (strain C); Species: Salmonella enterica subsp. enterica serovar Typhi

A; Note: this species has also been called Salmonella typhi

C; Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C; Accession: AB0539

R; Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A; Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A; Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi

A; Reference number: AB0502; MUID:21534947; PMID:11677608

A; Accession: AB0539

A; Status: preliminary

A; Molecule type: DNA

A; Residues: 1-843 <PAR>

A; Cross-references: GB:AL513382; PIDN:CAD08754.1; PID:gl6501575; GSPDB:GN00176

C; Genetics:

A; Gene: STY0324

Query Match 4.0%; Score 353.5; DB 2; Length 843;

Best Local Similarity 22.8%; Pred. No. 1.6e-11;

QY 1230 -----NRVSKQTDITDCELYR-----ETMLVNEVRNGEMI 1261
Db 1019 RERDLTGWMSLRKPEETWYGDORLTVTQOTRTQTVYQPSFPLRLRIETENGEOA 1078
QY 1262 RLLR-----TGTTIIAQ-----ORASKVLLTGTDGSOQS-----VILTSDKONLS 1300
Db 1079 KARHRSIAVLQEDBTGVTPLPAELAVMLGRLERELROGSVSEESQOQLAQCLTAEQMGQA 1138
QY 1301 QEA-----YSAYGKHSTANDASI-----LGYN 1323
Db 1139 LEAGVPIPERKHLHYHCDQGLPLGLISPGRETALTASYDWNLLSTSAQPIQQSLRFP 1198
QY 1324 GERADPVSUVTHLNGVRSYDPTLMRPHTPDPSLSPFG-AGGINPYSYCLGDPINRSDPSG 1382
Db 1199 GQOYDEESGLYNNRY--YDPLQGRYITQD---PIGLEGWNLYQVPL-NPIEHDPGLG 1252

RESULT 11
H91236
RhsH core protein with extension [imported] - Escherichia coli (strain O157:H7, substrain C; Species: Escherichia coli
C; Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C; Accession: H91236
R; Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A; Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic reference strains
A; Reference number: A99629; MUID: 21156231; PMID: 11258796
A; Accession: H91236
A; Status: preliminary
A; Molecule type: DNA
A; Residues: 1-1394 -HAV-
A; Cross-references: GB:BA000007; PIDN:BA038287.1; PID:gl3364340; GSPDB:GN00154
A; Experimental source: strain O157:H7, substrain RMD 0509952
C; Genetics:
A; Gene: Ecs4864
C; Superfamily: rhsF protein

Query March 3.8%; Score 341; DB 2; Length 1394;
Best Local Similarity 21.5%; Pred. No. 1.9e-10;
Matches 260; Conservative 156; Mismatches 404; Indels 392; Gaps 63;

QY 418 KETW-RSADNSYRSBITETTFDESIGNPLTKVI-----KDKTKIIS-----PS-----461
Db 218 RTQTFHRAAGFSGITGT-DGAGRHFELVLTQQAERARQQAISGGTEPAFPDT 276
QY 462 -----THWEYPPAGEVDNCPPEP---YGFT-----RFVK 488
Db 277 LPGYBYGRDNGIRLSAVMLTHDPEYP-----ENLPAALVRYGWTGPRGELAVVYDRSGK 331
QY 489 KIIQTPYDSEFK-----DDPKFIQVRYSLIGSQSHVTLKIEERHYSATQLLNSTL 539
Db 332 QVRSFYDDKYGRGVVAHRTGRPE--IRYRD---SGRVTQQLNPAFLSYT-----379
QY 540 FQNTDKSBLGRLKQTEC--TKGENKTYSVVHKFTYTKQDDTLQQSHSIITHDNFTIH 597
Db 380 YQVEKDRITITSLNRRVHLHQEGG-LKRVRVK-----EHADGSVT 421
QY 598 RSQVRSRYTRGFPDPTDKDITQMSYDKLGRLLITLNSGTPYANTLYDYELANLQDD 657
Db 422 QSQFDA--VGLRQAQDAAGRTYESPDVVTGLITRIT---TPDGRASAFY-----468
QY 658 NRPPFVITITDNGNQLRNEFDGAGRHVSQCLKSD-----GDGKFYTIHTQYDEQGRH 712
Db 469 NHHSQLTSATGPDGLREIREYDEWGLRIQETAPDGDITFYRDNPHSLDLPACATEDATGSR 528
QY 713 HTSTVSDYLTNGRQQTDPKXVHLSKMSYDNGVQANTHWSYGVSEKITYVPI-PLTATK 771
Db 529 KTWTSRY---GQLLSFTDCSGYVTRYDHFQGMATAVREBGLSQRYAYDSRGGLIAVK 585
QY 772 QIQSNNSNVQTKCVTVTPSPQPIQITLFDAGHLQSCHTLTRDGDWRVRKETAICQ- 830
Db 586 DTQGHETRYE-----YNAAGDLTT--VIAPDG-SRNGTQYDANGKA 623

QY 831 -CTI-----VOYDNNRVIOITLPPDGTIVNRKYAPFSTDTLITDIRVNGISLQOQTF- 881
Db 624 ICTTQGGTFRSEYDAAGRVIRLTSENGS-----HTTFRYDVLRLIOETGDFGRTORVH 678
QY 882 -DGLSRLTOSQDGRWVWYATYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASHI 940
Db 679 HULTGKLRSEDEGLV-----THWYD-EADRLTHRTVKGET 714
QY 941 TQOFSNPNVTGAL--LKAVAQCSLTPIY--PSGRLL--KMNENIMKKMSYLM-----988
Db 715 AERWOYDE-RGWLITDISHISEGHRVT-VHYGYDEKGRLTGERQTVHHQTEALLAQHETR 772
QY 989 ----TLRLENGYTDLTGTIQTIKLSRDTGHRVTKDSSIKTILNYDDLNRHIGSQVTDLAT 1045
Db 773 HAYNAQGLAN--RCIPDSLPAVENLTYG-----SGWLAGMKL 807
QY 1046 GHMLTTTVEF--DGLNREIGRKLCDSSGHTLIDIQOSWLKTCQLANRIVKLNGLVLRTEQY 1103
Db 808 GD--TFLVDFTDRDLRHLKTLRFRGYELTATVPAGQLSQSHLS-----LQYDRDY 857
QY 1104 SYDSRNLNOYKCDGACBPTDKYGHISVTONFTYDIYGNITACHTTFADGTEDHATFKPA 1163
Db 859 TWNDNGELIR-----ISSPRQ-----TRSYSDSGRLTGVHTTAA-----NLDIRIP 900
QY 1164 NPTDPC--QLTEVHHHTPD-----MPDN-----IRLKYDKAGRVINITD-----1200
Db 901 YATDPAGNRLPD-PELHDPDSLMSKPFNRIARDAHYLYRDRHGLRTEKTLIPGVIPT 959
QY 1201 NHGNTENFTYDGLRQNGQ-----GSVGYVDPLNRLVSOXTDLDCEL-----1244
Db 960 DDERTHRYHDSQHLRHLVHYTRQVEEPLVESRYLYDPLGRVAVKRVRRERDLTGWMSLS 1019
QY 1245 -----YV-----RETMLVNE-----VRNGEMIRLLR-----TGETIIAQRA-----1276
Db 1020 RKPQVYTWGDDRLTITQNDRTIQTQYQSGSTPLIRVTATGELAKTORRSLADALQ 1079
QY 1277 -----SKVLLTGTDSQOSVILTS-----1294
Db 1080 QSGGEDGSGVFPVPLVQVMDRLSEILADRVSESRRLASCGLTVAQMSQMDPVVTP 1139
QY 1295 -----DKQNL-----SOEA-----YSAGKHKSTAND-----ASILGYNGERADPV 1330
Db 1140 ARKHLVHCDHRGLPLALISKEGATEWCAEYDEWGNLNEENPHQLQOLIRLPQOQYDEE 1199
QY 1331 SGVTHLNGVRSYDPTLMRPHTPDPSLSPFG-AGGINPYSYCLGDPINRSDPSGHLSCWAW 1369
Db 1200 SGLYV--NRHRYVPLQGRYITQD---PIGLKGGWNLYTYPL-SPVNGMDPLGLYEFPSK 1253
QY 1390 TGIGMGAGLLL 1401
Db 1254 NIDDIGIFALAM 1265

RESULT 12
E65145
rhsB protein precursor - Escherichia coli (strain K-12)
C; Species: Escherichia coli
C; Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C; Accession: E65145; S47701; B36902; A30092; I54935
R; Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A; Title: The complete genome sequence of Escherichia coli K-12.
A; Reference number: A64720; MUID: 97426617; PMID: 9278503
A; Accession: E65145
A; Status: nucleic acid sequence not shown; translation not shown
A; Molecule type: DNA
A; Residues: 1-1411 <BLAT>
A; Cross-references: GB:AE000424; GB:U00096; NID:G9367230; PIDN:AACT6507.1; PID:gl789894
A; Experimental source: strain K-12, substrain MG1655
R; Plunkett, G.
Submitted to the EMBL Data Library, March 1994

A:Reference number: S47666
 A:Accession: S47701
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1129, 'Q', 1131-1411 <PLU>
 A:Cross-references: ENBL:U00039; NID:9466582; PIDN:AAB18457.1; PID:g466618
 A:Experimental source: strain K-12, substrain MG1655
 R:Zhao, S.; Sandt, C.H.; Feulner, G.; Vlazny, D.A.; Gray, J.A.; Hill, C.W.
 J. Bacteriol. 175, 2799-2808, 1993
 A:Title: Rhs elements of *Escherichia coli* K-12: complex composites of shared and unique
 A:Reference number: A36902; MUID:93259920; PMID:8387990
 A:Accession: B36902
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1397-1411 <ZHA>
 A>Note: sequence extracted from NCBI backbone (NCBIN:132073, NCBI:P.132075)
 R:Sadosky, A.B.; Davidson, A.; Lin, R.J.; Hill, C.W.
 J. Bacteriol. 171, 636-642, 1989
 A:Title: rhs gene family of *Escherichia coli* K-12.
 A:Reference number: A91901; MUID:89121333; PMID:2644231
 A:Accession: A30092
 A:Molecule type: DNA
 A:Residues: 1-100 <SAD>
 R:Feulner, G.; Gray, J.A.; Kirschmann, J.A.; Lehner, A.F.; Sadosky, A.B.; Vlazny, D.A.;
 J. Bacteriol. 172, 446-456, 1990
 A:Title: Structure of the rhsA locus from *Escherichia coli* K-12 and comparison of rhsA
 A:Reference number: 154935; MUID:90094253; PMID:2403547
 A:Accession: 154935
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1221-1411 <RS>
 A:Cross-references: GB:M29717; NID:g147622; PID:g147623
 C:Comment: the rhs core consist of two distinct parts: a large N-terminal core that is c
 C:Genetics:
 A:Gene: rhsB
 A:Map position: 77 min
 C:Superfamily: rhsF protein
 C:Keywords: transmembrane protein
 F:1-26/Domain: signal sequence #status predicted <SIG>
 F:27-1411/Product: rhsD protein #status predicted <MAN>
 F:28-55/Domain: transmembrane #status predicted <TM>

Query Match 3.8%; Score 340.5; DB 2; Length 1411;
 Best Local Similarity 20.4%; Pred. No. 2.1e-10;
 Matches 287; Conservative 172; Mismatches 459; Indels 490; Gaps 70;

QY	418	KTEW-RGADNSYRSEITETTFDESIGNPLTKVI-----KDKTKTKIIS-----PS-----	461
DB	218	RQTHREAGEFSGEITGVTDGAWHFRVLVLTQARAEARQQAISGTEFSAPDPT	276
QY	462	-----THWYYPAGEVDNCPPEP---YGFT-----	
DB	277	LPFYTEYGRDNGIRLSAVWLTHDPEP-----ENLPAALVRYGWTGPRGELAVVYDRSGK	331
QY	489	KIIOPPYDFEFK-----DDPEKIQYRSLIGSQSHVTLKIBERHYSATQLLNSTL	539
DB	332	QVRSFYDDKYRGWVAHHTGREP--IRVYD---SDGRVTBQLNAPAGLSYI-----	379
QY	540	FOYNTDKSELGRLLKQTEC--TKGENKTVSVVHKFTYTKQDDTLQOSHSHIITHDPTTH	597
DB	380	YQYKDRITITDLSLRREVLTQGEAG-LKRVVKK-----EHADGSVT	421
QY	598	RSQVRSYTGSLFSDTKDKIVTQMSYDKLGRLLTRTLNSGTVPANTLYDYELNQLDD	657
DB	422	QSQFPA--VGLRAQTDAAGTTEYSDDVTGLITRIT---TQDGRASAFYNNHQQ---	473
QY	658	NRPPVITTTDVNGNQLNEPDGAGRHVSQCLKSD-----GSGKYFTHTQYDEQGRH	712
DB	474	-----LTSATGPDGLELRREYDELGRLIQETAPDGDITRYRDNPHSDLPFCATEDATGR	528
QY	713	HTSYSDYLITNGRQOTPDKVHLSMSKSYDNWGOIANTHWSYGVSEKITVDPT--TLTATK	771
DB	529	KMTWSRY--GQLLSFTDCSGYTRYDHRFQMTAVHREGLSQRAYVDSRGQLIAVK	585

QY	772	QLOSNNNVQTKREVITYTPSQOPIQITLDFEAGHLQ-----SCHTLTRCDWRVRKE	824
DB	586	DTQGHETRYE-----YNIAGDUTAVIAPDGSRRNGTQYDANGKAVIT	626
QY	825	TDALCOCTIYOYNNRNVIOITLPDGTIVNRKAPFSTDLITDIRVNGISLQQOFTF--D	882
DB	627	TQG-GLTRSMEDYDAAGRVIRLTSENGS-----HTTFRYDVLDRLIQETGFGDGTQRYHHD	680
QY	883	GLSLTQSCDGRVWATYTSAGNDQCPSTVITPDGFIHYVQOPELDDAVLQVANSNEITQ	942
DB	681	LTGKLIRSEDEGLV-----THWHYD-EADRLTHRTVRKETAETAE	716
QY	943	QFSYNPVTGAL--LKAVAEGQSLLPIYV---PSGRL--KMNINDMKMSYLW-----	988
DB	717	RWOYDE-RGWLTDISHISEGHRVA-VHYRYDEKGRLTGERQTVHHPTQETALLWQHETRA	774
QY	989	TLRGLNGYTDLTGTTICKISRDTGHRVTOIKDSSIKTILNYDDLNHRHIGSVQTDLATOH	1047
DB	775	YNAOGLAN--RCIPDSDLPVAVENITYG-----SGVLAKWKLGD-----	809
QY	1048	MLTTTVEF--DGLNREIGRKLCDSSGHTLIDQOSWLKTOQLANRIVKLVGLQTEQKSY	1105
DB	810	-TPLEVTRDLHRETLR-----SFGRYELTAYTPAGLOQSO--HLNSLLS-DRDYTW	859
QY	1106	DSRNLNQKCDGACPTDKYGHISVTQNTFYDIYGNITACHTTFADGTEDHATKFNAP	1165
DB	860	NDNGELIR-----ISSRQ-----TRSYSTTGRLTGVHTTAA-----NLDIRIPVA	902
QY	1166	TDPG--QLTEVHTHDPD-----MPDN-----IRLYDKAGRVNITD-----NH	1202
DB	903	TDPAGNELPD-PELHPDSTLSMPDNRIARDAHVLVYDRHGRLEKTEKTLIPEGVIRTD	961
QY	1203	GNTENFYDILGRLOQC-----GSVGYDPLNRLVSKQTDPLDCEL-----	1244
DB	962	ERTHYHYDSQHLRVHVTQYEEPLVESGRYLYDPGRVVRVRRRDLTGWMSLSRK	1021
QY	1245	---YY-----RETMLVNE-----VRNGEMIRLLR---TGETIIAQORA-----	1276
DB	1022	PQVTWYMGDGRLLTIQNDRSRIQTIYQPGSFPLIRVETATGELAKTQRRSLADALQOS	1081
QY	1277	-----SKVLLTGTDSQSVILTSQKNLSQEA-----	1303
DB	1082	GGEDGSVFPFVPLVQMLDESEILLA---DRVSEESRWLASCLGTVQMNQNDPVVT	1138
QY	1304	-----YSAYGKHKSTAND---ASILGYNGERADP	1329
DB	1139	PARKIHLVCHDRGLPLALISTEGATAWCAEYDEWGNLLNEENPHQLQLIRLPQOQYDE	1198
QY	1330	VSGYTHLNGYRSYDPTLRFHTPDSLSBPG-AGGINPYSYCLGDDINSDPSGHLWSQA	1388
DB	1199	ESGLYY--NHRYYDPLQGRYITQD---PIGLKGNWLYGYQL-NFISIDPLGLSMWED	1252
QY	1389	WTGIMGIAGLLLTIATGNGMAIAAAGIAAATSTTALAFGALSVTSDITSIVSGALE	1448
DB	1253	-----AKSG--ACTNGLCGTISA-----MIGDPKDFSDIDSTAY	1283
QY	1449	DASPKASSILGWVSMGMGAAGLAESAIGKGTKLATHGAFADGENALLKSTSESSRIKW	1508
DB	1284	DALNKINS-----QSICDEKPEA-----	1301
QY	1509	GVTRSLDREIVRNEEGOVIKDHSRGYTDNFMKGEO-----AILVHGDKDGFYHTEG	1561
DB	1302	-----GLICKNSGRYFSTAPNRRGERKGSYPFNSPCPNCTEKVSAYHTHG	1346
QY	1562	NKNGK--GPYTRHTPEQLVDYLKDNKI	1587
DB	1347	ADSHGEYDEIFSGKDEKIVK-SKONNI	1373

RESULT 13

H64780

rhsD protein precursor - *Escherichia coli* (strain K-12)

C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence revision 17-Sep-1997 #text_change 01-Mar-2002
 C:Accession: J54780; J50625; B30092; I69401; S16026
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 .A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: J54780
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1426 <BLAT>
 A:Cross-references: GB:AB000156; GB:U00096; NID:G1786705; PIDN:AAC73599.1; PID:G1786706;
 A:Experimental source: strain K-12, substrain Mgl655
 R:Sadosky, A.B.; Gray, J.A.; Hill, C.W.
 Nucleic Acids Res. 19, 7177-7183, 1991
 A:Title: The rhd-E subfamily of Escherichia coli K-12.
 A:Reference number: J50625; MUID:92115567; PMID:1766878
 A:Accession: J50625
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-378 'A', 380-1166 'G', 1168-1426 <SAD>
 A:Cross-references: EMBL:X60999; NID:G42732; PID:G42733
 A:Experimental source: strain K-12
 R:Sadosky, A.B.; Davidson, A.; Lin, R.J.; Hill, C.W.
 J. Bacteriol. 171, 636-642, 1989
 A:Title: The rhd gene family of Escherichia coli K-12.
 A:Reference number: A91901; MUID:89123133; PMID:2644231
 A:Accession: B30092
 A:Molecule type: DNA
 A:Residues: 1-100 <SAD>
 A:Cross-references: GB:J04224; GB:J04224; NID:G147646; PIDN:AAA24542.1; PID:G147649
 R:Feulner, G.; Gray, J.A.; Kirschmann, J.A.; Lehner, A.F.; Sadosky, A.B.; Vlazny, D.A.;
 J. Bacteriol. 172, 446-456, 1990
 A:Title: Structure of the rhdA locus from Escherichia coli K-12 and comparison of rhdA
 A:Reference number: I54935; MUID:90094253; PMID:2403547
 A:Accession: I69401
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1232-1426 <RES>
 A:Cross-references: GB:M29719; NID:G147644; PIDN:AAA24541.1; PID:G147645
 C:Comment: the rhd core consist of two distinct parts: a large N-terminal core that is c
 C:Genetics:
 A:Gene: rhd
 C:Superfamily: rhdF protein
 C:Keywords: transmembrane protein
 F:1-26/Domain: signal sequence #status predicted <SIG>
 F:27-1426/Product: rhd protein #status predicted <MAT>
 F:28-55/Domain: transmembrane #status predicted <TM>
 Query Match 3.8%; Score 339; DB 2; Length 1426;
 Best Local Similarity 20.2%; Pred. No. 2.5e-10;
 Matches 257; Conservative 150; Mismatches 409; Indels 458; Gaps 60;
 QY 572 KFTYTKQDDTLQOQSHSTTHDFTIHRQVRSRYTGRLEFSDTDTKDIVTQMSYDKLGRLL 631
 DB 317 RYITTEAGELL-----AVYDSNTQVAFYDAQHPGMVAHRYAGPEMRYDYDTGRVV 372
 QY 632 TRTNGSTPVANTLYDYELNNLQDNRPPFVITTD-----EQGRHITSYSDYL 721
 DB 373 EQLNPAG-----LSYRY-----LYEQDR-----ITVTDLSNRREVLHTEGGAGLKRVMKKE 418
 QY 669 -VNGNQLNRFEDGAGRHVSQCLKSDGDKFTHIQDYD-----EQGRHITSYSDYL 721
 DB 419 LADGSTRSYDAAGRLTAQ--TDAAGRTEYGLNVSGDITDITPDGRTKTFYND-- 474
 QY 722 TNGRQQT---DPPKVLHSMKSYDNNWQGIANTHWSYGVSEKITVDPITLTATKQLQSNRN 778
 DB 475 --GNQLTAVVSPD--GLESREYDEPGLV-----SE 502
 QY 779 NVOTGKEVT--TYTPSQOPIQITLFDGAGLQCHLITRDGWDNRVKETDAICQCTIYQVD 837
 DB 503 TSKSGETVRYRDAHSELFPATTTDATG---STROMTWSRYGQLLAFTDCSGYQTRYEYD 559

RESULT 14

E85509

hypothetical protein Z0268 [imported] - Escherichia coli (strain O157:H7, substrain EDL9

C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E85509
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamianos, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85509
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1404 <STO>
A:Cross-references: GB:AE005174; NID:gl2512977; PIDN:AAG54537.1; GSPDB:GN00145; UWGP:Z02
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z0268
C:Superfamily: rhaF protein

Query Match 3.88; Score 336.5; DB 2; Length 1404;
Best Local Similarity 20.08; Pred. No. 3.4e-10;
Matches 252; Conservative 130; Mismatches 373; Indels 505; Gaps 58;

Qy 376 YHL-LTSECKQNGYIQTAYAYAIIGHNFDSPQSQFOLPKT-KTETWRSADNSYRSEI 433
Db 245 FHLVLTQAQAEAFVRQRAATSLSSPAGPR--SASSSLVFPDPLPAGTEYGADNGIRLEA 302
Qy 434 TETTFDES-----GNPLTKVKKDKTKKILSPSTHWEYPPAGEV---DNCPPPYGF 483
Db 303 VMLTHDPAYPDPLPAAPLARYT-----YTASGELRAVYDRSGTVQVRG 345
Qy 484 TRFKKLIQTPDSE-----FKDDPEKFTQYRYSLIGSQSHVTLKIEERHYGATOL 534
Db 346 A-----YDAEHAGRVVAHYAGRPES--RYRYDDTG-----RVTEL 379
Qy 535 LN-----STLFQYNTDKSELGRLLKQTEC--TKGENKTVSVVHKFTYTKQDDTLQOHSI 588
Db 380 VNPGLDYRPEYGDRTVITDLSNRREVLYTEGEGG-LKEWVK-----422
Qy 589 TTHDNFTIHSOVSRYTGRFSDTDKDIVTQMSYDKGLRLTRLNSGTPYANTLYD 648
Db 423 -EHADGSGITSE-----YDEAGRLAQT-----444
Qy 649 YELNLQDNRPPFVITTDVNGNLNREFDGRHVSQCLKDSG-----DGKFTY 700
Db 445 -----DAAGRRTEYSLHMASGAVTAVTGPDGR--T 472
Qy 701 IHTQYDEQGRHHTSYSDVLTNGRQOTDPDKVHLMSKSYDNWQGIANTHWSYGVSEKI 760
Db 473 VR-YGYSQRQVTSVTPDGLRSREYDEKRLAAETSRGE-----TTRYSD-----520
Qy 761 TVDPITLTATKQLOSGNNGVQTEKVTYTPSQOPIQITLFDGAGHLQSCHTLRGDWR 820
Db 521 --DP-----ASELPTGIQDATGSTKQM-----AMSR 544
Qy 821 ---VRKETDAIGCTIYQYDYNRVITLTPDGTIVNRKYAPSTDTLIDRVNGISLG 877
Db 545 YGQLLTFTDCSGYTRYEYDRYGOQIAVHREGLSTYSSYNP-----RG 588
Qy 878 QOTFDGLSRITQSQGGRVWATYSAGNDQCPSTVITPDQGFHYQYQPELDDA---VLQ 934
Db 589 Q-----LVSQDAQGRREYVSAAGDL--TAIVAPDGRSRIQY-----DAWGKAVS 634
Qy 935 VASNEITQFSPNPVNTGALLKVAEGOSLTPPIYPSGRLLKMNINMKMSYLTWLEGLE 994
Db 635 TTQGLTRSMGYDAAGRITVLTNENGQSSTFRYPDVLRLTEQRFQDGTORY-----686
Qy 995 NGYTDLTGTTI-----QKISRTHGRVTOIKDSS-- 1022
Db 687 ---HYDLGKLTQSEDEGLVTLWHYDASDRITHRVNCGDPAEQWYDEHGLWLTLSHTSEG 744
Qy 1023 --IKTLLNYDNLNHHISQ---VTOLATGHMI-----1049
Db 745 HRVSVHYGYDDKGLRTGERQTVENPETGEMLWEHETGHAYSEQGLATRQSPDGLPPVEWL 804

Qy 1050 -----TTTVEF--DOLNREIGRKLCDSSGHTLIDQOQSWLKTLQOLANRLVKL 1093
Db 805 TYSGYLAGMKLGCTPLVEYMRDLHRETARSF--GGEAYELATAWNTSGQURSRLNL 861
Qy 1094 NGVLQTEQSYDSRNLNQKCDGAECPDKYGHSTVQNFYDIYGNITACHTTTPADG 1153
Db 862 P---QLDRDYDNDNGQL--IRISGPQ-----ESREYRSDTGRLTGVHTTAANL 906
Qy 1154 TED--HATFKFANP-TDPCQLTEVHHTHPD-----MPDNRL-----KYDKAGRVIN 1197
Db 907 DIDIPYATDPAGNRPDP-----ELHDDSTLTAWPDN-RIAEADAHVYVRYDEYGRLA 958
Qy 1198 ITD-----NHGNTENTYDTLGR-----QNGQGSV---YGYDPL-----1229
Db 959 KTDRIPEGVIRMDERTHHYHDSQHLRVFHTRIHQGEPOVESRYLYDPLGRRTGKRV 1018
Qy 1230 -----NRLVSOKTDTLDCELYR-----ETMLNVEVRNGEMI 1261
Db 1019 RERDLTGWMSLSRKPETWYGMWDGRLTIVTQOTRIQTVYQPGSFTPLRIETENGEQA 1078
Qy 1262 RLLR-----TGETIIAQ-----QRASKVLLTGTDSQOS-----1289
Db 1079 KARHSLAEVLQEDTGTVTPAEALVLMGLRLERLRQGSVSESOQWLACGLTAEQMAAQ 1138
Qy 1290 -----VILTSDKONLSQEAAYSAYGK---HKSTANDASILGYN 1323
Db 1139 LEAEVYPERKHLHYCHDRGLPLALISPEGETAWQGEYDEWGNLLGETSACHLQQSLRLP 1198
Qy 1324 GERADPVGTVHLNGVRSYDPTLMRPHTPDLSLSPFG-AGGINPYSYCLGDPINRSPSG 1382
Db 1199 CQOYDEESGLYNNRY--YDPLQGRYITOD---PIGLEGWNLYQPL-NPIEHIDPLG 1252

RESULT 15
C65159
rhaA protein precursor - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: C65159; S47814; I69402
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A:Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97425617; PMID:9278503
A:Accession: C65159
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1377 <BLAT>
A:Cross-references: GB:AE000437; GB:U00096; NID:G2367249; PIDN:AAC76617.1; PID:gl790020;
A:Experimental source: strain K-12, substrain MGL1655
R:Plunkett, G.
submitted to the EMBL Data Library, March 1994
A:Reference number: S47666
A:Accession: S47814
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-864, 'M', 866-1377 <PLU>
A:Cross-references: EMBL:U00039; NID:946582; PIDN:AA818570.1; PID:g466731
A:Experimental source: strain K-12, substrain MGL1655
R:Feulner, G.; Gray, J.A.; Kirschmann, J.A.; Jenner, A.F.; Sadosky, A.B.; Vlazny, D.A.;
J. Bacteriol. 172, 446-456, 1990
A:Title: Structure of the rhaA locus from Escherichia coli K-12 and comparison of rhaA w
A:Reference number: I54935; MUID:90094253; PMID:2403547
A:Accession: I69402
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-864, 'M', 866-1377 <RES>
A:Cross-references: GB:M29716; NID:gl47613; PID:gl47614
A:Experimental source: strain K-12
C:Comment: the rha core consist of two distinct parts: a large N-terminal core that is co
C:Genetics:
A:Gene: rhaA

C;Superfamily: rhsF protein
C;Keywords: transmembrane protein
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-1377/Product: rhsD protein #status predicted <MAT>
F;28-55/Domain: transmembrane #status predicted <TM>

Query Match 3.7%; Score 331.5; DB 2; Length 1377;
Best local Similarity 20.8%; Pred. No. 6.2e-10;
Matches 288; Conservative 170; Mismatches 450; Indels 477; Gaps 75;

QY 211 LFPGQKEGY-RTELPLNRLQNLGNHPLNPLTWSFGYTGKNGILGWITMTAP 269
DB 126 LFQGE-DGYSSESLLVR--GGVAKLDEGRLAALWQ----- 160

QY 270 GGLKETVNYG-----NNQG-----HHFQSANL--PVLPIVLMKQVPGAQ- 310
DB 161 -ALPEELRLSPHYRLATNSPQGPWLLGWCERPEADVLPAPLPPYRLVLTGLVDFRGT 219

QY 311 PAIOAEVSYTSHNVGGSGNGIWNKLDNLGLMTEYNYGSTESRYKDKGEGHDIQVRIE 370
DB 220 QTHREAGFSGEITGVTDGAWRH-----FRVL-----TQAQR----- 255

QY 371 RTYNNVHLLTSECKQNGYIQTETAYAIIGHNFDSPSQFQPKT---KTETWRSADN 427
DB 256 -----AEARQQ-----AISG--GTEPSAF--PDTLPGYTEYGR--DN 237

QY 428 SYRSEITETTFDESNGPLTKVIKDKTKQIIISPSTHWEYPPAGEVDNCPPEP---YGET 484
DB 288 GIRLSAVMLTHDPE-----YP-----ENLFAAPLVRYGWT 317

QY 485 -----RFVKIITQPYDSEFK-----DDPEKFIQYRYSLSQSOSHVTLKIE 525
DB 318 PRGELAVVYDRSGKQVSFYDYDKYGRMVVHRHTGRPE--IRYRD---SDGRVTEQLN 372

QY 526 ERIYSATOLLNSTLTFQNTDKSELGRLLKQTEC--TKGENGKTYSVVHKFTYTKQDDTIQ 593
DB 373 PAGLSYT-----YQEKDRITITDSDLREVLHTQGEAG-LKRVVVK----- 413

QY 584 QSHSITTHDNFTIHRSVRSGRYGRFSDPTDKDIVTQMSYDKLGRLLTTLNSGTPYAN 643
DB 414 -----EHADGVTQSQFDA--VRLRAQTDAGRTETSPDVVTGLITRIT---TPDGR 462

QY 644 TLTYDEYELNLDNRPFPVITTTDNGNQLRNEFPDAGRHVSQCLKDS-----GDGKF 698
DB 463 ASAFYNNHNQ-----LTSATGPDGLELREYDELGRLIQETAPDGDITRYRYDNPH 514

QY 699 YTIHTQYQDEGRHRTSYDYLTLNGRQQTDPKVLHLSKSYDNWGOIANTHWSYGVSE 758
DB 515 SDLPATEDATGSRKNTWSRY---GQLSFTQSGVYTRYDHDRCQMTAVHREGLSQ 571

QY 759 KITVDPI-TLTATKQLQSNNNVQGTKEVTTYTPSQPIQITLDFEAGHLQ-----SC 810
DB 572 YRAVDSRGQLIAVKDTQGHETRYE-----YNIAGDLTAVIAPDGSR 612

QY 811 HTLTRDGNDRVKETDAIGCTIYQYDNNYRVITQITLPGDTIVNRKYAPSPDITLDIR 870
DB 613 NGTQYDANGKAVRTQG-GTRSMEXYDAAGRVILTSNGS-----HTFRYDVLDRLLQ 666

QY 871 VNGISLGCQTF--DGLSRLTQSDGGRVWAYTYSAGNDQCPSTVITPDGOFIHYOYQPEL 928
DB 667 ETGFDGRTQRYHHDLTGKLIRSEDEGLV-----THWHYD-EA 702

QY 929 DDAVLQVASNEITQQFSYNPVTGAL--LKAVASQSLTPIYY---PSGRLL--KQENINDM 981
DB 703 DRLHTRVKGTAERWQYDE-RGWLTDISHISEGHRVA-VHRYRYDEKGRLTGERQTVHPH 760

QY 982 KQWSYLW-----TLRGLNGYTLDTGTLQKISRDTGHRVTQIKDSSIKTLLNYDDLN 1033
DB 761 QTEALLQWHEHTRAYNAQGLAN--RCIPDPLPAVEMLTYG-----SGVLAKGLGD-- 809

QY 1034 RHIGSVQVTLATGHMLTTTVEF--DGLNREIGRKLCDSSGHTLIDQOSWLKTOQLANRIV 1091
DB 810 -----TPLVEYTRDLRHRETLR-----SPGRVELTTAVTPAQQLQSQ-- 846

QY 1092 KLINGVLQRTQYSDYSRNLNQYKCDGAECPTDKYGHISIVTQNFYDIYGNITACHTTFA 1151
DB 847 HLNSLLS--DRDYTWNDNGELIR-----ISSPRQ-----TRSYSYSTTGLTGWHITAA 893

QY 1152 DGTEDHATFKFANPTDPC--QLTEVHHTHPD-----MPDN-----IRLKYDKAGRVIN 1197
DB 894 -----NLDIRIPYATDPAGNRLPD--PELHPDSTLSMPDNDRIARDAHYLYRYDRHGRUTE 947

QY 1198 ITD-----NHGNTENFTYDTLGRQLNQO-----GSVGYDPLNRLYSQKTDIT 1239
DB 948 KTDLIPEGVIRTDDBERTHRVHYDSQHRVLVHVTRTYEPEPLVESRYLYDPLGRRVAKRVWR 1007

QY 1240 LDCEL-----YY-----RETMVNE-----VRNGEMIRLLR-----TGFTI 1270
DB 1008 RERDLTGWMSLSRKFPQVTWYCWGDGRLTIQNDTRIQTIIYQPGSFTEPLIRVETATGELA 1067

QY 1271 LAQORA-----SKVLLTGTDSQQSVIL-----TS 1294
DB 1068 KTORSLADALQOQSGGEGGSGVPPVVLVQMLDRLSEILLADRVSEBSRRRLASCGLTVE 1127

QY 1295 DKQN-----LSQEA-----YSAYGKHKSAND---AS 1318
DB 1128 QWQNMDFVYTPARKIHLHYCHDRGLPLALISKEGTTWCAEYDEWGNLLNEENPHQLQQ 1187

QY 1319 ILGYNGERADPVSGVTHLNGYRSDPTLMSEFHTPDSLSRPF--AGGINPYSYCLGDPINR 1377
DB 1188 LIRLPQOQYDEESGLY--NRHRYDPLQGRYITQD---PIGLKGMNFYQYPL-NPVTN 1241

QY 1378 SDPSG 1382
DB 1242 TDPLG 1246

Search completed: July 3, 2004, 05:47:37
Job time : 38 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2004, 05:42:23 ; Search time 19 Seconds
(without alignments)
4584.914 Million cell updates/sec

Title: US-09-889-874A-23

Perfect score: 8879

Sequence: 1 VYIKFLKPRITMSDNEF.....PRKILGRTEKTVKPTFRP 1673

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	432	4.9	2334	1 WAPA_BACSU	Q07833 bacillus su
2	342	3.9	1397	1 RHSD_ECOLI	P16918 escherichia
3	341.5	3.8	1411	1 RHSE_ECOLI	P16917 escherichia
4	339	3.8	1426	1 RHSD_ECOLI	P16919 escherichia
5	331.5	3.7	1377	1 RHSE_ECOLI	P16916 escherichia
6	220.5	2.5	1183	1 CNA_STAAR	Q53654 staphylococ
7	216.5	2.4	2894	1 YD96_METJA	Q58791 methanococ
8	206	2.3	1902	1 P2P_LACLC	P15293 lactococcus
9	205.5	2.3	1902	1 P3P_LACLC	P15292 lactococcus
10	201	2.3	1902	1 P1P_LACLC	P16271 lactococcus
11	200.5	2.3	1902	1 P2P_LACPA	Q02470 lactobacill
12	198	2.2	2660	1 YEEJ_ECO57	Q04877 escherichia
13	196	2.2	2867	1 RBP2_PLAVB	Q00799 plasmodium
14	188.5	2.1	682	1 RHSE_ECOLI	P24211 escherichia
15	186.5	2.1	2003	1 YDBA_ECOLI	P33666 escherichia
16	185.5	2.1	2358	1 YEEJ_ECOLI	P76347 escherichia
17	185	2.1	1861	1 APU_THETU	P38536 t amylpull
18	183	2.1	3381	1 PGCV_BOVIN	P81282 bos taurus
19	181	2.0	1228	1 SLAP_BACST	P35825 bacillus st
20	180.5	2.0	1577	1 YL3A_PROMI	P16466 proteus mir
21	178.5	2.0	3178	1 YS89_CABEL	Q09624 caenorhabdi
22	174.5	2.0	1803	1 YL3_YEAST	P47024 saccharomyc
23	173.5	2.0	1257	1 CCAA_BACTU	Q45754 bacillus th
24	173	1.9	1829	1 FRPC_NEIMB	Q93975 neisseria m
25	172	1.9	1829	1 FRPC_NEIMC	P55127 neisseria m
26	171.5	1.9	1592	1 GTF2_STRDO	P27470 streptococc
27	170.5	1.9	3097	1 CADN_DROVE	O15943 drosophila
28	169.5	1.9	2314	1 PTP2_HUMAN	P23471 homo sapien
29	168	1.9	1744	1 TANA_XENLA	Q01550 xenopus lae
30	165	1.9	828	1 PMFC_PROMI	P53514 proteus mir
31	163	1.8	1698	1 4I_DROME	Q09890 drosophila
32	163	1.8	3358	1 PGCV_MOUSE	Q62059 mus musculu
33	161	1.8	1953	1 BIGA_SALTY	P25927 salmonella

34 160.5 1.8 1018 1 FNBA_STAAR
35 160.5 1.8 1560 1 TENN_MOUSE
36 160.5 1.8 2021 1 OMPA_RICCN
37 160 1.8 1569 1 YFUA_ECOLI
38 159.5 1.8 1523 1 DPOL_THEFM
39 159 1.8 1409 1 HAP1_HAEIN
40 159 1.8 1723 1 PM20_CHLPN
41 158.5 1.8 1656 1 OMPB_RICJA
42 158 1.8 1812 1 SUIS_SUNMU
43 157 1.8 2145 1 CYAA_PODAN
44 156.5 1.8 1271 1 Y338_MYCGE
45 156.5 1.8 1609 1 FIG2_YEAST

ALIGNMENTS

RESULT 1
WAPA_BACSU
ID WAPA_BACSU STANDARD; PRT; 2334 AA.
AC Q07833;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Wall-associated protein precursor.
GN WAPA OR N17G OR BSU9230.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]_TaxID=1423;
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=93302506; PubMed=8316082;
RA Foster S.J.;
RT "Molecular analysis of three major wall-associated proteins of
RT Bacillus subtilis 168: evidence for processing of the product of a
RT gene encoding a 258 kDa precursor two-domain ligand-binding
RT protein."
RL Mol. Microbiol. 8:299-310(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=95319088; PubMed=7704263;
RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
RT genome containing the hut and wapa loci."
RL Microbiology 141:337-343(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=97124196; PubMed=8969509;
RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
RA Miwa Y., Fujita Y.;
RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
RT containing the lic and cel loci, and creation of a 177 kb contig
RT covering the gnt-eachy region."
RL Microbiology 142:3113-3123(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bartero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Denizot K.D., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian F., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Chim S.Y., Glaser P., Goffeau A., Goughly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,

P14738 staphylococ
Q80271 mus musculu
Q52657 rickettsia
P52143 escherichia
P74918 thermococcu
P44596 haemophilus
Q92812 chlamydia p
O62653 r outer mem
O62653 suncus muri
Q01513 podospora a
P47580 mycoplasma
P25653 saccharomyc

RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Paro V., Pohl J.M., Portetelle D., Porwolik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpestra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vaesarotti A.,
RA Viari A., Wambutt R., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis";
RL Nature 390:249-256(1997).
CC -!- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
CC MOTILITY, SECRETION OR DIFFERENTIATION.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
CC INTO THE MEDIUM.
CC -!- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE
CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
CC MOTIF REPEATED 31 TIMES.
CC -!- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
CC SIMILARITY TO THE REPEAT IN E. COLI RHS GROUP OF PROTEINS (RHS-A-D).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sb-sib.ch).
CC -----
CC EMBL; L05634; AAA22883.1; -
CC EMBL; D31856; BAA06556.1; -
CC EMBL; D29985; BAA06260.1; -
CC EMBL; D83026; BAA11683.1; -
CC EMBL; Z99124; CAB15959.1; -
CC PIR; S32920; S32920.
CC Subtilisin; BG10797; wapa.
CC InterPro; IPR003305; CBM_CenC.
CC InterPro; IPR004530; YD.
CC Pfam; PF02018; CBM_4_9; 1.
CC Pfam; PF05593; RHS_repeat; 14.
CC TIGRFAMs; TIGR01643; YD_repeat_2x; 17.
CC Cell wall; Repeat; Signal; Complete proteome.
CC OR 32 (POTENTIAL).
CC SIGNAL 1 28
CC CHAIN 29 2334
CC DOMAIN 504 869
CC REPEAT 504 605
CC REPEAT 504 605
CC REPEAT 636 736
CC REPEAT 769 869
CC DOMAIN 1021 2139
CC REPEAT 1021 1040
CC REPEAT 1042 1061
CC REPEAT 1063 1082
CC REPEAT 1083 1102
CC REPEAT 1109 1128
CC REPEAT 1129 1148
CC REPEAT 1150 1169
CC REPEAT 1174 1193
CC REPEAT 1199 1218
CC REPEAT 1219 1238
CC REPEAT 1646 1665
CC REPEAT 1667 1686
CC REPEAT 1690 1709
CC REPEAT 1711 1730

FT REPEAT 1732 1751
FT REPEAT 1753 1772
FT REPEAT 1795 1814
FT REPEAT 1820 1839
FT REPEAT 1840 1859
FT REPEAT 1861 1880
FT REPEAT 1887 1906
FT REPEAT 1908 1927
FT REPEAT 1929 1948
FT REPEAT 1969 1982
FT REPEAT 1983 2002
FT REPEAT 2008 2027
FT REPEAT 2028 2047
FT REPEAT 2051 2070
FT REPEAT 2071 2090
FT REPEAT 2093 2112
FT REPEAT 2120 2139
SQ SEQUENCE 2334 AA; 258329 MW; B75138CCD278BAA3 CRC64;
Query Match 4.9%; Score 432; DB 1; Length 2334;
Best Local Similarity 20.4%; Pred. No. 3.2e-15;
Matches 358; Conservative 226; Mismatches 613; Indels 560; Gaps 80;
QY 44 IQITLGHIVG-NGNL-----GPTLPLTILSYSPLNKTDIGFGIGFNGLSVYDR 90
DB 896 IDPSQLNGATGNVIVNEEDLSIDGRGPGGLGRTYNSLSDSSDHLFGQGW-----YADA 950
QY 91 KNSLLSLSTGNYKVIETDKTVLQOKKLDN-----LRFKDLKENCYRIIHKSGDIEVL 145
DB 951 ETSVISTDGGAMY-IDEDATHTRTKADGTQPTGVILELTETADQFLTKTD-----1004
QY 146 TGFNNNAFDLKVPKLL-----NPAGHAIYDWN-----FEATQPLNRYTDOLDGH-- 192
DB 1005 ---QTNNAIFNKKGKLQKVVDGHNNAVTYNDKQNLTAITDASGRKLTFTYDB-NGHVT 1060
QY 193 -----DIPLLMLEY--QGLI-----KTILTFPG 214
DB 1061 SITGPKNKVTYSYENDLLKKVTDGTVTSYDYDSEGLVKQYSANSTAKPVFTY-- 1118
QY 215 QKEGYRTELRLNQLNSIHNS-----LGNENPLTWSFGYTPICKNGILGOWITSMT 267
DB 1119 QYSGHRLKALNAKKTIVYSYDADKKTLLMQNGRKVKQYGYNEAGNP-----IQVID 1172
QY 268 APGLKKTWN--YSNNN-----QGHFPOSANLPVLPYVTLMKQVPGAG 309
DB 1173 DAEGLKITNTKYGNNVVEDVDNDVGTGKATESYQYDKDGN-----VTSVKDAYGT- 1225
QY 310 QPATAEYSYTSNHYVGGSGNGIWNKLDNLYGLMTEYNGSTESRYKDKEGHDQIVRI 369
DB 1226 -----ETYEYNNKNDV-----TKMKDTEGNTVTDIAYDGLDAVSETDQSGKSSAAV 1271
QY 370 ERTYNNYHLLTSECKQONGYIQTETAYAIIGHNFDSPQSQFQPKTKTETWRSADNSY 429
DB 1272 YDKYGNQIQSSKDLASNNILK-----DGSFEAQKSGWNLTASKD-----1311
QY 430 RSEITETTFDESNGPLTKVIKDKTKLIISPS-----THWEYPPAGVDNCPPEPYGFR 485
DB 1312 RKIS-VIADKSG-----VLGSKALEVLQSSTAGTDHGYSSATQVLEPNTYTLTG 1365
QY 486 FVKIIL-----QTPYDSEFKDDPEKFIQY---RYSLIGSQSHVTLKIBERHYSATOLLNS-- 537
DB 1366 KIKTDLAKSRAVFNIDLRDQKQKLOIWHNEYSALACKNDWT-----KQITFTTPAKAG 1421
QY 538 ---TLFQVNTDKSELGRL-LQTECTKGKGTYSVVKHTYTKQDDTLQOS-HSITHD 592
DB 1422 AVVYMEVDHKDKGKAWFDEVLKEGEVSSYPVQNSFTSATENWNVNGASVDSEE 1481
QY 593 NFTIHRSQVRSYTGRLFSDDTDKDIQVTKSYDKLGRLLTRT-----L 635
DB 1482 CFNDDVSLKAARTSASQAGSVTKQTVVLGQSANDKPVLTLTGMSKASSVKFTDKDYSL 1541
QY 636 NSGTFYANTLTYYDEL---NNLQDNRPPFVI----- 664

Db 1542 QANVTYADGSTGIYNAKPSGTQGNRAAVVVPKTPINKYKINXKIDISILFQKSAITGVWFDI 1601
Qy 665 -----TTTVANGNLNRNFDGAGRHVSQCLKSDGDGKGYFIHTHQVDEQGRHHT 714
Db 1602 RLIEGSLITKSTYSDNSGNYVTKBEDLGYATS-----TDYDTGK-----KTSETDAKGEKTT 1654
Qy 715 STY--SDYLNTRGQQTDPKRVHLSMSKSYDNWG-QIANT-----HWSGVSEKI-- 760
Db 1655 YTDQADQLNMTLSNGTSILH-----SYDKENVSXKIRAGADQTYFEFVDMGKLVK 1709
Qy 761 TVDPITLTATKQSNNSNNVQT-----GKEVTTYTPSQQFIQITLFDGAGHLSCHTLTRD 816
Db 1710 TTDPLGNVLASEYDANSNLTKTISPNQNEV-----SLSYD 1744
Qy 817 GWRVRKETDAICQCTIYQYDYNRVNIQITLPGTIVNRKYAPSPSTDTLITDIRVNGISL 876
Db 1745 GTRVRSKSYNGTEKIFITYDKNGN-----ETSVNKEQN-----TT 1781
Qy 877 GQOTFDGLSLTQSQDGRWVATYSAGNDQCSTVITPDQFIHQYQPELDDAVLQVA 936
Db 1782 KKEFDKRNRLTETLDEGGSCWTYPSDSKLTFF-----SWIH-----G 1821
Qy 937 SNEITQFNSVNVGTGALLKAVASQSLTPYIPSPGRLKMNINDMKMSVYLTWLRGLNG 996
Db 1822 DQKGTNQFTYN-----KLDQMTKMDSTSSSYFDYDEN- 1954
Qy 997 YTDLTGTIQKISRDTGSRVTOIKDSSIKTTLAYDNLNRHIGSOVTLATGHMLTITVTFD 1056
Db 1855 -----GNVQ-----TFITGNGGTSFSDERLNVSLHIGDKNGDILTESYEV- 1898
Qy 1057 GLNREIGRKLCDSGHGLTDIQQWMLKTOQLANRIVKLVNGVLQRTQYSDYSDNRNLNOYKC 1116
Db 1899 -----DANGNRNTINSS-----ASGKV-----QVEYKLNQLVK-- 1927
Qy 1117 DGAECPTDKYHSIVTONFNYDLYGNITACHTTFADGTED--HATKFNPTDPCOLTEV 1174
Db 1928 -----ETHEDGTVI---EYIDGFGNKTV-TTIKQSSKTVNASNIMW-----QJTKV 1973
Qy 1175 H-----HTHP-DMPDNI-----RLKYDKAGRVNIINIDNHG 1203
Db 1974 NDESISYDKNGNRSTDGKFTYTDAEDNLAVTKGDKGKPFATYKYDKGNRIQKTVN-G 2032
Qy 1204 NTFNTYDTLGRLONGGSGVGVDPNLNLSQKTDILDCELY-YRETMVLVNRNGEMTR 1262
Db 2033 KVTNYFYDVG-----DSLNLVYETDADNNVTKSYTYGD-----SQGLLS 2070
Qy 1263 LLRTGETIIAQORASKVLLLTGTDSSQSVILTSKONLSQAYSGYKGR-HKSTANDA---S 1318
Db 2071 YTENGKKYFHYNAAGDIIAISDSGTGV-----AKQYDANGNPKTEASDEVKON 2122
Qy 1319 ILGVNGERADPVSGVTHLNGYRSYDPTLMRFHT--PDSLSPFGAGGGINPYCYCLGDPIN 1376
Db 2123 RYRYAGYQYDEETGLYYLMARY--YEPRNGVFLSLDPDPSGDSGLDQNGYAYGNPNVYM 2180
Qy 1377 RSDPSGHLSCQAWTIGCGTAGLLLTATGMAIARAAG---GIARAIASTSTTALAFA 1432
Db 2181 NVDPDGH--W-VMLVNVNAGFA-----AYDGYKAYKSGKWKGAANAASNFPQKIFKG 2231
Qy 1433 LSVTSDITSIVSGALEDASPASKASSILGWVSNMG-----GAAGLAESAIKGCTKLATHLGA 1488
Db 2232 ASRAYKTK-----KAVKITGHTRHGLNQSIGRNG-----GRGVNLRKALNA- 2273
Qy 1489 AEDGENALLKSTSSRIKGVTRSLDRE--IVRNEEGOVIKDHSRGYTNDFMGKGEQAI 1546
Db 2274 -----VRSPPKVIQPNKATYGVGKATVNLNRKGVITAYG-----SS 2312
Qy 1547 LVHGDKDGFVHTEGK 1563
Db 2313 RAKSKVFTHTHGK 2329

RESULT 2

RHSC_ECOLI

ID AC RHSC_ECOLI STANDARD; PRT; 1397 AA.
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE RHSC protein precursor.
GN RHSC OR B0700.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=93259920; PubMed=8387990;
RA Zhao S., Sandt C.H., Feulner G., Vlazny D.A., Gray J.A., Hill C.W.;
RT "Rhs elements of Escherichia coli K-12: complex composites of shared
RL and unique components that have different evolutionary histories.";
RN J. Bacteriol. 175:2799-2808(1993).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kiragawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RL corresponding to the 12.7-28.0 min region on the linkage map.";
RN DNA Res. 3:137-155(1996).
RP [4]
RP SEQUENCE OF 1-100 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=89123133; PubMed=2644231;
RA Sadosky A.B., Davidson A., Lin R.J., Hill C.W.;
RT "Rhs gene family of Escherichia coli K-12.";
RL J. Bacteriol. 171:636-642(1989).
RN [5]
RP SEQUENCE OF 1221-1397 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90094253; PubMed=2403547;
RA Feulner G., Gray J.A., Kirschman J.A., Lehner A.F., Sadosky A.B.,
RA Vlazny D.A., Zhang J., Zhao S., Hill C.W.;
RT "Structure of the rhaA locus from Escherichia coli K-12 and
RL comparison of rhaA with other members of the rhs multigene family.";
RN J. Bacteriol. 172:446-456(1990).
RP [6]
RP REVIEW.
RX MEDLINE=95020608; PubMed=7934896;
RA Hill C.W., Sandt C.H., Vlazny D.A.;
RT "Rhs elements of Escherichia coli: a family of genetic composites
RL each encoding a large mosaic protein.";
RN Mol. Microbiol. 12:865-871(1994).
CC -!- FUNCTION: Rhs elements have a nonessential function. They may play
CC an important role in the natural ecology of the cell.
CC -!- DOMAIN: Each rhs appears to consist of a highly conserved 141 kDa
CC amino fragment followed by a highly divergent carboxy terminus.
CC -!- SIMILARITY: BELONGS TO THE RHS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE RhsB protein precursor.
GN RhsB OR B3482.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_taxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=9325920; PubMed=9387990;
RA Zhao S., Sandt C.H., Feulner G., Vlazny D.A., Gray J.A., Hill C.W.;
RT "Rhs elements of Escherichia coli K-12: complex composites of shared
RT and unique components that have different evolutionary histories";
RL J. Bacteriol. 175:2799-2808(1993).
RN [2]
RP REVISION TO 405.
RA Hill C.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94318500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes";
RL Nucleic Acids Res. 22:2576-2586(1994).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12";
RL Science 277:1453-1474(1997).
RN [5]
RP SEQUENCE OF 1-100 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=8912133; PubMed=2644231;
RA Sadosky A.B., Davidson A., Lin R.J., Hill C.W.;
RT "Rhs gene family of Escherichia coli K-12";
RL J. Bacteriol. 171:636-642(1989).
RN [6]
RP SEQUENCE OF 1221-1411 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90094253; PubMed=2403547;
RA Feulner G., Gray J.A., Kirschman J.A., Lehner A.P., Sadosky A.B.,
RA Vlazny D.A., Zhang J., Zhao S., Hill C.W.;
RT "Structure of the rhaA locus from Escherichia coli K-12 and
RT comparison of rhaA with other members of the rhaA multigene family";
RL J. Bacteriol. 172:446-456(1990).
RN [7]
RP REVIEW
RX MEDLINE=95020608; PubMed=7934896;
RA Hill C.W., Sandt C.H., Vlazny D.A.;
RT "Rhs elements of Escherichia coli: a family of genetic composites
RT each encoding a large mosaic protein";
RL Mol. Microbiol. 12:865-871(1994).
CC CC -!- FUNCTION: Rhs elements have a nonessential function. They may play
CC an important role in the natural ecology of the cell.
CC -!- DOMAIN: Each rhaA appears to consist of a highly conserved 141 kDa
CC amino fragment followed by a highly divergent carboxy terminus.
CC -!- SIMILARITY: BELONGS TO THE Rhs FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; L02370; AAC61883.1; -
DR EMBL; U00039; AAB18457.1; -
DR EMBL; AE000424; AAC76507.1; -
DR PIR; E65145; E65145.
DR EcoGene; EGI0847; rhsB.
DR InterPro; IPR001826; Rhs.
DR InterPro; IPR006530; YD.
DR Pfam; PF03527; Rhs; 1.
DR Pfam; PF05593; Rhs_repeat; 13.
DR PRINTS; PRO0394; RhsPROTEIN.
DR TIGRFAMs; TIGR01643; YD repeat 2x; 15.
KW Signal; Multigene family; Transmembrane; Repeat; Complete proteome.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 1411 RhsB PROTEIN.
FT TRANSMEM 28 55 POTENTIAL.
FT DOMAIN 330 1186 28 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 330 352 1.
FT REPEAT 353 374 2.
FT REPEAT 375 417 3.
FT REPEAT 418 438 4.
FT REPEAT 439 460 5.
FT REPEAT 461 481 6.
FT REPEAT 482 502 7.
FT REPEAT 503 525 8.
FT REPEAT 526 546 9.
FT REPEAT 547 567 10.
FT REPEAT 568 588 11.
FT REPEAT 589 609 12.
FT REPEAT 610 629 13.
FT REPEAT 630 650 14.
FT REPEAT 651 671 15.
FT REPEAT 672 691 16.
FT REPEAT 692 711 17.
FT REPEAT 712 734 18.
FT REPEAT 735 758 19.
FT REPEAT 808 828 20.
FT REPEAT 829 850 21.
FT REPEAT 851 871 22.
FT REPEAT 872 894 23.
FT REPEAT 895 930 24.
FT REPEAT 931 959 25.
FT REPEAT 960 984 26.
FT REPEAT 985 1019 27.
FT REPEAT 1162 1186 28.
FT CONFLICT 1130 1130 Q -> K (IN REF. 4).
SQ SEQUENCE 1411 AA; 159394 MW; 567AC4EE713D9E07 CRC64;

Query Match 3.8%; Score 341.5; DB 1; Length 1411;
Best Local Similarity 20.4%; Pred. No. 1.3e-10; Indels 484; Gaps 70;
Matches 287; Conservative 172; Mismatches 462;

QY 418 KETW-RSADNSYRSEITETTFESGNPLTKVI-----KDKTKIIS-----PS----- 461
DB 218 RTCTFHRAAGFSGEITGVT-DGAWRHFLVLTTCQAQAEAEARCAISGGTEPSAFPTD 276
QY 462 -----TWYYPAGEVDNCPPEP---YGET-----RFLK 488
DB 277 LPQYTEYGDNGCRISAVWLTHDPEYP-----ENIPAPLVRYGWTFRGELAVVYDRSGK 331
QY 489 KIIQTYDSEFK-----DDPEKFIQYRSLIGSQSHVTKIEERHYSATQLINSL 539
DB 332 QVRSFYDDKYGRVVAHRTGEP-----IYRYD---SDGRVTEQLNPAGLSVT----- 379
QY 540 FQNTDKSELGRLLKQTEC---TKGNGKTYSVVHKFTYTKQDQTLQOQSHIITHNFTIH 597
DB 380 YQVEKDRITITDLSRREVLHTQGEAG-LKRYVK-----EHADGSVT 421
QY 598 RQVRSRYGRFLPSDDTKDVIQMSYDKLGRLLTTLNSGTPYANTLYDYELANLQDD 657
DB 422 QSQFDA--VGRRAQTDAAGRTEVSPDVVTGLITRIT---TPDGRASAFYNNHNO--- 473
QY 658 NRPPFVITTTDVGNGQLRNEFDGAGHVSQCLKSD-----GDGKPYTHITQYDEQGRH 712

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; X60999; CAA43314.1; --
DR EMBL; AE000156; AAC73593.1; --
DR EMBL; U82664; AAE40251.1; --
DR EMBL; AF044500; AAC32467.1; --
DR EMBL; M21764; AAA24542.1; --
DR EMBL; M29719; AAA24541.1; --
DR PIR; H64780; H64780.
DR EcoGene; EG10849; rbsD.
DR InterPro; IPR001826; RBS.
DR InterPro; IPR006530; YD.
DR Pfam; PF03527; RBS; 1.
DR Pfam; PF05593; RBS; repeat; 11.
DR PRINTS; PR00394; RBS; PROTEIN.
DR TIGRFAMs; TIGR01643; YD repeat 2x; 16.
KW Signal; Multigene family; Transmembrane; Repeat; Complete proteome.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 1426 RBS; PROTEIN.
FT TRANSMEM 28 55 POTENTIAL.
FT CONFLICT 379 379 G -> A (IN REF. 1).
FT CONFLICT 1167 1167 A -> G (IN REF. 1).
SQ SEQUENCE 1426 AA; 159724 MW; 122D697AA49BC2F CRC64;

Query Match 3.8%; Score 339; DB 1; Length 1426;
Best Local Similarity 20.2%; Pred. No. 1.8e-10;
Matches 257; Conservative 150; Mismatches 409; Indels 458; Gaps 60;
QY 572 KFTYTKDDDTQQQSHITTHDNFIHRSQVRSYTGELFSDTDPKDIVTQMSYDKLGRLL 631
DB 317 RYVTEAGELL---AVYDRSNTQVRAFTYDAQHPGRVAHVRAGREMYRVDIGRV 372
QY 632 TRTNSGTPYANTITYYELNLODDNRPPVITTTD----- 668
DB 373 EQLNPAG-----LSYRY-----LYEQDR---IVTDSLNRREVLHTEGGAGLKRVRVKE 418
QY 669 -VNGNQLRNEFDGAGRVSOCLKSDGDGKFTYTHTOQYD-----EQGRHHSTYSYDL 721
DB 419 LADSVTRSGVDAAGRLTAQ--TPAAGRTEYGLNVVSGDITDITPDGRETFFYND-- 474
QY 722 TNGRQQT---DPPDKVHLSMSKSYDNWQCIANTHWSYGVSEKITVDPIITATKQLQSN 778
DB 475 --GNQLTAVVSPD--GLESRREYDEPGRV-----SE 502
QY 779 NVOTGKEVT--TYTPSQOPIQITLDEAGHLQSCHTLTRDGDWRVRKETDAIGQCTTYQYD 837
DB 503 TSRSGETVRVRYDDAHSELPAITTDATG---STROMTWSRYGQLLAFTDCSGYQTRYED 559
QY 838 NYARVQITLFDGTVNKRKAPSTDLITDIRVNGISLQQQTFDGLSRLTSQDQ-GRV 896
DB 560 RFGQM-----TAVHRE-----EGISL--YRYDNRGLTSVKDAQRE 595
QY 897 WAYTYSACNQCPSTVITPDGQFIHYOQ----- 925
DB 596 TRYENAAAGL--TAVITPDGNRSETQYDAGKAVSTTQGLRSMEDYDAAGRVISLTNE 653
QY 926 -----PELDDAVLQVANSNETTQOFSYNPVTGALLKAVARGQSITPIYPSGLKME 976
DB 654 NGSHVSFSDALDELVOQGGFDGRTQRYHYD-ITGKLTQSEDEGLVILWYDESDRTHR 712
QY 977 NINDMKQVSLWTLRGLNGVYDLTGFIQKISDRTHGRVTQIKDSS-----IKTLLNYDDL 1032
DB 713 TVNGEPAEQWQY-----DGHGWLTDIHLSEGHVRVAHYGYDDK 751
QY 1033 NRHIG--SQVTDLATGHML----- 1049
DB 752 GLTGECCQVETPETGELLWQHETKAYNEQGLANRVTPDPLPPVWLTGYSGYLGMKL 911

QY 1050 ---TTTVEF--DGLNREIGRKLCDSSGH--TLDIQQSWLKTQQLANR----- 1089
DB 812 GGTPLVEYTRDLRHLRETVRSFGSMAGSNAAYELTSTYTPAGQLSQHLSLVYDRDYGWS 871
QY 1090 ---IVKLANGVLQRTQYSVDSNRNLNQKCDGAC-----PTDKYGHSI----- 1130
DB 872 DNGDLVRISGPRQ--TREYGYSATGRLESVRTLAPDLDIRIPYATDPAGNRLDPPELHPDS 930
QY 1131 -----VTQN-----PTYDIYGNITACHTTTPADG---TEDHATKFFANPTDFCOLTEV 1174
DB 931 TLTWPFNDRIAEADAHVYVRHDEYGRULTKTRIPAGVIRTDDBERT----- 975
QY 1175 HHTHPDMPDNIRLKYDKXGRVINITD-NHGN---TENFTYDTLGR----- 1215
DB 976 HHYH-----YDSQHLRVFTYRIQGEPLVESRYDPLGRMAKRVRRERDLTG 1025
QY 1216 ---LQNGQSVYGYDPLNRLVSQKTDITLDCELYR----- 1247
DB 1026 WMSLSRKPEVTWYMGDQ-DRLTTVQDITTRITQTVYEPGFTPLIRVETENGEREKAQRS 1084
QY 1248 ---ETMLVNEVRNG-----EMIRLL-RGETIIIA-----QORASKV 1279
DB 1085 LAETLQOEGSENGHGVVFPFAELVRLDLRLEESIRADRVSSSESRAMLQOGLTVEQLARQV 1144
QY 1280 LLTGTDSQOS-----VILTSDKONLSQEA--YSAYGKHKSTANDASILGYN--- 1323
DB 1145 EPEYTPARKAHLVCHDRGLPLALISEDGNTAWSAEYDEWGNQLEENPHV--YQPYL 1202
QY 1324 -GERADPVSGVTHLNGVRSYDFTLMRPHTPDSLSPFG--AGGINPYSYCLGDPINRSDPS 1381
DB 1203 PQQHQDEESGLY--NRHRYDPLQGRYITQD---PMGLKGGWNLVYQYPL-NPLQOIDS 1256
QY 1382 GHLSMQAWTGIGMGIAGLLTATGMAIAAAGGIAAATASTTALAFGALSVTSDITS 1441
DB 1257 GLL--QTWDDARSG-----ACTGV-----CGVLSRIIGPSKFDSTADAALD----- 1296
QY 1442 IVSGALEDASPKASSILGWVSMGAAGLAESAIKXGTTKLATHLGAFAEDGENALLKSTS 1501
DB 1297 ---ALKETQNRS-----LCNDMEYSGIVCKDTNG-----KYFASKAETDNLK-K 1336
QY 1502 ESSRLKMGVTRSLDRREIVNEREGQVYKHSRQYTDNEMWKGFOAILVHGDK--DGFLLHT 1559
DB 1337 ESYPLURKCKPTQDTRVAAYHTHG---ADSHGDIYDVEFFSSSDKNLVRKONNLEAFYLAT 1393
QY 1560 EGNKH---NGKGPY 1570
DB 1394 PDGRPEALNNKGEY 1407

RESULT 5

RHSA_ECOLI
ID RHSA_ECOLI STANDARD; PRT; 1377 AA.
AC P16916;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE RhsA protein precursor.
GN RHSA OR B3593.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID:562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90094253; PubMed=2403547;
RA Feulner G., Gray J.A., Kirschman J.A., Lehner A.F., Sadosky A.B.,
RA Vlazny D.A., Zhang J., Zhao S., Hill C.W.,
RT "Structure of the rhaA locus from Escherichia coli K-12 and
RT comparison of rhaA with other members of the rha multigene family."
RL J. Bacteriol. 172:446-456(1990).
RN [2]

SEQUENCE FROM N.A.
 RP STRAIN=K12 / MG1555;
 RC MEDLINE=94316500; PubMed=8041620;
 RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
 RL region from 76.0 to 81.5 minutes.";
 RN Nucleic Acids Res. 22:2576-2586(1994).
 [3]
 RP REVIEW.
 RX MEDLINE=95020608; PubMed=7934896;
 RA Hill C.W., Sandt C.H., Vlazny D.A.;
 RT "Rhs elements of Escherichia coli: a family of genetic composites
 RL each encoding a large mosaic protein.";
 CC Mol. Microbiol. 12:865-871(1994).
 CC -!- FUNCTION: Rhs elements have a nonessential function. They may play
 CC an important role in the natural ecology of the cell.
 CC -!- DOMAIN: Each rhs appears to consist of a highly conserved 141 kDa
 CC amino fragment followed by a highly divergent carboxy terminus.
 CC -!- SIMILARITY: BELONGS TO THE Rhs FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L19044; AAC95065.1; -;
 DR EMBL: U00039; AAB18570.1; -;
 DR EMBL: AE000437; AAC76617.1; -;
 DR PIR: C65159; C65159.
 DR EcoGene: EG10846; rhsA.
 DR InterPro: IPR001826; Rhs.
 DR InterPro: IPR006530; YD.
 DR Pfam: PF03527; Rhs; 1.
 DR PRINTS: PR00394; Rhs repeat; 12.
 DR TIGRFAMs: TIGR01643; YD repeat_2x; 15.
 DR Signal: Multigene family; Transmembrane; Repeat; Complete proteome.
 KW SIGNAL: 1 26
 FT CHAIN 27 1377
 FT TRANSMEM 28 55
 FT DOMAIN 330 1186
 FT REPEAT 330 352
 FT REPEAT 353 374
 FT REPEAT 375 417
 FT REPEAT 418 438
 FT REPEAT 439 460
 FT REPEAT 461 481
 FT REPEAT 482 502
 FT REPEAT 503 525
 FT REPEAT 526 546
 FT REPEAT 547 567
 FT REPEAT 568 588
 FT REPEAT 589 609
 FT REPEAT 610 629
 FT REPEAT 630 650
 FT REPEAT 651 671
 FT REPEAT 672 691
 FT REPEAT 692 711
 FT REPEAT 712 734
 FT REPEAT 735 758
 FT REPEAT 759 782
 FT REPEAT 783 808
 FT REPEAT 809 828
 FT REPEAT 829 851
 FT REPEAT 851 871
 FT REPEAT 872 894
 FT REPEAT 895 930
 FT REPEAT 931 959
 FT REPEAT 960 984
 FT REPEAT 985 1019
 FT REPEAT 1162 1186
 FT REPEAT 1377 AA; 156320 MM; 21ACA989E74200FE CRC64;
 SQ SEQUENCE

Query Match 3.7%; Score 331.5; DB 1; Length 1377;
 Best Local Similarity 20.8%; Pred. No. 4.3e-10;
 Matches 288; Conservative 170; Mismatches 450; Indels 477; Gaps 75;
 211 LPPGQKEGY-RTEFLNRLQNLNHNFLSGNENPLTWSFGYTPGKNGILQWNTSMTAP 269
 126 LPPGE-DGYSRSESLVLR--GGVAKLDEGHRLAALMQ----- 160
 270 GGLKETVNY-----NNNQ-----HHFQSANL-----PVLVYVTLMKQVPGAGQ- 310
 161 ALPEELRLSPHYLATNSPQGPWMLLGCERVPDEVLPAFLPYRVLTGLVDRFGRT 219
 311 PAIQAEYSVTSNHYVGGSGNGIWNKLDNLGLMTEYNGSTESRRYKDKEGHQIVRIE 370
 220 QTFHREAAEGFSGEITGVDGAWRH-----FRLVL-----TTQAR----- 255
 371 RTNNYHLLTSECKQNGYIOTTETAYYAIIGHNFDSPQSFQPKT---KTETWRSADN 427
 256 -----ABEARQQ-----AISG-----GTEPSAF--PDTLPGYTEYGR--DN 287
 428 SVRSITFTTDESNGPLTKVKKTKQKIISPSTHWEYPPAGEVDCPEP---YGF 484
 288 GIRLSAVMLTHDPE-----Y-----ENLPAAPLVRVGT 317
 485 -----RFVKXIIQTPYDSBFK-----DDPEKFIQYRYSLIGSQSHVTKIE 525
 318 PRGELAVVYDRSGKQVRSFTYDDKYGRWVAHRHTGRPE--IRYRD---SDGRVTEQLN 372
 526 ERHYSATQLLNSLTFOYNTDKSELGRLLKQTEC--TKGNGKTSYVVKHFTYTKQDDTLQ 583
 373 PAGLSVT-----YQEKDRITITDSLRREVLHTQGEAG-LKRVVK----- 413
 584 QSHSITTHDNFTIHRSQVRSRYTGRFSDTDTKDITVQMSYDKLGRLLTRTLNSGTPYAN 643
 414 -----EHADGSVTSQFPA--VGLRAQTDAAGTTEVSVDVWVGLTRIT--TPDGR 462
 644 TLTYDYELNLODDNRPFFVITTTDNGNQLRNEFDGAGRHSVQCLKDS-----GDKGF 698
 463 ASAFYNNHNQ-----LTSATPGDLRLREYDELGRLIQETAPDGLTRVRYDNPH 514
 699 YTHIQQDEQGRHHTSYSDYLTNGROQTDPKVHLSMSKSYDNWGOIANTHWSYGVSE 758
 515 SLDPCATEDATGSRKMTWSRY---GQLLSFTDCSGYTRVDHDFGQMTAVHREGLSQ 571
 759 KITVDFI-TLTATKQLQNSNNVQTKVTTTPSQQPIQITLFEAGHLQ-----SC 810
 572 YRAVDSRGQLIAVKDTQGHETRYE-----YNIAGDLTAVIAPDGSR 612
 811 HLTRDGDWRVRKETDAIGCCTIYOYDNVNRVIGITLPGTIVNRKYPFSTDTLITDIR 870
 613 NGTYDAMGKAVRTTQG-GLTRSMYDAAGRVIRITSENGS-----HTFRVDVLDRLIQ 666
 871 VNGISLGQQTFF--DGLSRITQSDGGRVWAYTYSAGNQCPCSTVITPDQFIHYQBEL 928
 667 ETGFDGRTQRYHDLTKLIRSEDEGLV-----THMYD-EA 702
 929 DRAVLQVAGNEITQOFSYNPVTGAL--LKAVAEGSLTPIVY---PSGRL--KMNINDM 981
 703 DRLTHRTVKGETAERWQYDE-RGWLTDISHISEGHRVA-VHYRYDEKRLTGERGTVHP 760
 982 KKMXYLW-----TLRGLNGYDITGCTIKISRDTHGRVTQIKDSSIKTLLNYDDLIN 1033
 761 QTEALLWQHETHRAYNAOGLAN--RCIPDSLPAVWELTVG-----SGYLAGMKLGD-- 809
 1034 RHIGSQVTDLATGCHMLTTTVEP--DGLNREIGRKLDCSSGHSLDLCQSLKLTQQLANRIV 1091
 810 -----TPLVEYTRDLRHLRETLR-----SFGRYELTATYPAGQLQSQ-- 846
 1092 KLVGLQRTQYSDGRNRLNQVKDGAECPTDKYGHSHSVITQNTFYDIYGNITACHTTFA 1151
 847 HLNSLLS-DRDYTWNDNGELIR-----ISSPRQ-----TRSYSTTGRLTGVHTTAA 893

```
QY 1152 DGTEDHATKFNATDPC--OLTEVHTHPD-----MPDN-----IRLKYDKAGRVN 1197
DB 894 -----NLDIRIPYATDPAGNLPD--PELHPDSTLSMPDNRIARDAHYLYRDSHGLTE 947
QY 1198 LTD-----NHGNTENFYDILGRLOQC-----GSVGYDPLNRLVSQKTD 1239
DB 948 KIDLPEGVITDDERTHRYHDSQRLVHVTRTQYEELVESRYLYDPLGRVAKVR 1007
QY 1240 LDCEL-----YY-----RETMLVNE-----VRNGEMIRLLR-----TGCTI 1270
DB 1008 RERDLTGWMSLSRKPVQWYMGDGRLLTIQNDRTIQTIQPSFTPLRVERATGELA 1067
QY 1271 IAQRA-----SKULLTDSQCVIL-----TS 1294
DB 1068 KTORSLADALQSGGEGDGSVFPFVLVQMLDRLESEILADRVSESRRLWASGLTVE 1127
QY 1295 DKQN-----LSQEA-----YSAYCKHKSTAND---AS 1318
DB 1128 QMOMQMDPVYTPARKIHLXCHDRGLPLALISKEGTTEWCAYDEWGNLNEENPHOLOQ 1187
QY 1319 ILVNGERADVPVGVTHLNGYRSYDPTLMFHTPDSLSPEF--AGGINPYSYCLGDPINR 1377
DB 1188 LIRLPGQYDRESGLYY--NEHRYVDPLQGRYITQD---PIGLKGGWNFQYPL-NDVTN 1241
QY 1378 SDPSG 1382
DB 1242 TDPLG 1246

RESULT 6
CNA_STA00 STANDARD; PRT; 1183 AA.
AC Q53654,
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen adhesin precursor.
CN CNA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FDA 574;
RX MEDLINE=92165839; PubMed=1311320;
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RA Lindberg M., Hoesek M.;
RT "Molecular characterization and expression of a gene encoding a
RT Staphylococcus aureus collagen adhesin."
RL J. Biol. Chem. 267:4766-4772(1992).
RN [2]
RP ERRATUM.
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RA Lindberg M., Hoesek M.;
RA J. Biol. Chem. 269:11672-11672(1994).
RN [3]
RP COLLAGEN-BINDING DOMAIN.
RC STRAIN=FDA 574;
RX MEDLINE=94032261; PubMed=8218209;
RA Patti J.M., Boles J.O., Hoesek M.;
RT "Identification and biochemical characterization of the ligand
RT binding domain of the collagen adhesin from Staphylococcus aureus."
RL Biochemistry 32:11428-11435(1993).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
RX MEDLINE=97475225; PubMed=9334749;
RA Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,
RA Moore D., Jin L., Schneider A., DeLucas L.J., Hoesek M.,
RA Narayana S.V.L.;
RT "Structure of the collagen-binding domain from a Staphylococcus
RT aureus adhesin."
RL Nat. Struct. Biol. 4:833-838(1997).
CC -1- FUNCTION: MEDIATES ATTACHMENT OF STAPHYOCOCCAL CELLS TO
```

```
CC COLLAGEN-CONTAINING SUBSTRATA.
CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M81736; AAA20874.1; -
DR PDB; 1AMX; 24-JUN-98.
DR PDB; 1D20; 27-SEP-00.
DR PDB; 1D2P; 27-SEP-00.
DR InterPro; IPR008966; Adhes_bact.
DR InterPro; IPR008454; Cna_B_unit.
DR InterPro; IPR008970; Cna_B_unit.
DR InterPro; IPR008456; Collagen_bind.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna_B; 7.
DR Pfam; PF05737; Collagen_bind; 1.
DR TIGRfam; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; Repeat; Signal; 3D-structure.
KW Cell wall; Peptidoglycan-anchor; Repeat; Signal; 3D-structure.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 1154 COLLAGEN ADHESIN.
FT PROPEP 1155 1183 REMOVED BY SORTASE (POTENTIAL).
FT DOMAIN 151 318 COLLAGEN-BINDING.
FT DOMAIN 533 1093 3 X 187 AA APPROXIMATE TANDEM REPEATS.
FT DOMAIN 1093 1157 LYS/PRO-RICH (CELL WALL-SPANNING).
FT REPEAT 533 719 B1.
FT REPEAT 720 906 B2.
FT REPEAT 907 1093 B3.
FT SITE 1151 1155 LPXTG SORTING SIGNAL (POTENTIAL).
FT MOD_RES 1154 1154 AMIDE-LINKED TO CELL WALL (POTENTIAL).
FT STRAND 174 179
FT TURN 182 183
FT TURN 185 186
FT STRAND 187 194
FT TURN 196 197
FT STRAND 201 201
FT STRAND 205 211
FT STRAND 215 228
FT TURN 229 230
FT STRAND 232 234
FT HELIX 239 246
FT TURN 248 249
FT STRAND 251 255
FT TURN 256 259
FT STRAND 260 265
FT HELIX 267 270
FT TURN 271 272
FT STRAND 273 283
FT TURN 286 287
FT STRAND 290 299
FT STRAND 301 301
FT TURN 302 303
FT STRAND 307 311
FT STRAND 314 317
SQ SEQUENCE 1183 AA; 133066 MW; B6A1CC072E575D76 CRC64;

Query Match 2.5%; Score 220.5; DB 1; Length 1183;
Best Local Similarity 19.1%; Pred. No. 0.00039;
Matches 206; Conservative 159; Mismatches 405; Indels 311; Gaps 56;
QY 299 VTLKQVFCAGQ---PAIQAEYSYTSNHYVGGSGNGIWNKLDNLYGLMTYNYGTSR 355
DB 205 ITRKQIQGGQDLSTLNINVTGTHSNYSGQS-----AITDFKAFPGSK 251
QY 356 RYDKEGHGDIVRIERTYNNYHLLTSECK-----QQNGYIQTETAYTAIG----- 402
```


Db	252	ITVDNTKNTIDVTIPOQYSGVNSFSINYYKTIWNEQKQFVNSQ-AWYQEHGKEEVNGK	310
Qy	403	-----HNFDSPQSQFQPKTKTTWRSANDSYSEITETTFBDSGNPLTKYVKDKKTQK	456
Db	311	SNFHTVHNINAGIEGTGKELVKLQ-KDRTKAPIANVKFKLSKD-GSVVKD--NQK	366
Qy	457	IISPSTHWEYPPAGEVDCPEPVGFTRFKKIIQTPYDSEPKDDPEKFIQYRSL--I	514
Db	367	BI-----EITDANGIANIKALPSG--DYILKEIAPRYTFDKXE-----YPTFMKDT	414
Qy	515	GSQSHVTL-----KIE-----BRHYSATOLLNSTLF-----QYNT-----DKSELGR	551
Db	415	DNQGFYTIENAKALEKTKDVSQAQKWEQTKVKTPTIYFKLYKQDDNQNTTPVDKABIKK	474
Qy	552	LLKQTECK-----GNGKTYSVVHKETYTKQDDTLQOSSH-----ITHDNFTI	596
Db	475	LEDGT--TKVTWSNLPENDKNGKAIKYLKVEVNAQGEDTTPGTYKKENGVLVNTTEKPI	532
Qy	597	HRQVRSRYTGRLFSDTDTKD--IVTQMSYDKLG--RLTRTLNSGTPYANTILYDYEL	651
Db	533	ETTSISGE---KWDDXNDQDGKPEKVSVNLLANGKVKTLDTVSET-----NWKYEF	583
Qy	652	NNL-----QDNRPFFVIT-----TTDVGNGQLRNEFDGAGRHSQCCLKSDGDKGYTI	701
Db	584	KDLPYDEGKIEYTVTEHDVKDYTDINGTTITNKYT-PGETSATVTWKWDDNN-----	637
Qy	702	HTQQYDEQGRHHT-----STYDYLTLNGRQOTDPDKVHLSMSKSYDNMGQIANTHWSYGS	757
Db	638	NQGKPEIKVELYODGRATGK-----TAILNESNNW-----HTWTGLD	678
Qy	738	E-----KITVDPITLTKOLQSNVNVOTGEVIT--YTPS-----	792
Db	679	EKAQGOQVKYVE--ELAKVAGYTHVDNDMGNLIVTNKYTPETTSISGEKWDDKDNQ	736
Qy	793	---COPIQITLFDKAGHLSCHTLTRDQWDRVKEITDAIGCCTIYOYDNNRVQITLP	848
Db	737	DGKRPEKVSNNLADGEKVKTLDTVSETNWKYFKD-----LPKYDEGKIEYIVTE	788
Qy	849	D-----GTIVNRKYAPFSDTLI-----TDIRV-----NGISLQO-	878
Db	789	DVHKDYTTDINGTTITNKYTPGETSATVTKNWDDNNODGKRPEIKVELYQDGKATGT	848
Qy	879	-----QTFDGLSRLTQSO-----DGRVWATYSAGNDQCPSTVITPDGQFIH	921
Db	849	AILNESNNWHTWTGLDEKAGQVKYVTEELTKVGYTHVDNDMGNLIVT-----	901
Qy	922	YOYQPELQDAVLQVANSNITQFQSYNPVPTGALLKAVAEQSLTPIYPSPGELKMNNDM	981
Db	902	NKYTPETTSISGEKWDDKDNQDGKRP-EKVSNNLANGEKVKTLDTVSETNWKYEFKD	960
Qy	982	-----KMSYLWTLRGLENGWTDLTGTIOKISRTHGRVTOIKDSSIKITLNYDDLNRH	1035
Db	961	PKYDEGKIEYTVTEHDVKDYTDINGT-TINKYTPG-----ETSATVTKNWDDNNQ	1013
Qy	1036	IGSQVTLATGHLMTTVEF--DGLNREIGRKLCDSSGHTLIDIQOS--WLKTOQLANRIV	1091
Db	1014	DGKRPEI-----KVELYQDG-----KATKTALESNNWHTWTGLDEKA	1055
Qy	1092	KXNGVLQRTQYVSYSRNLNQQKCGAECPTDKYGHSLVTONFT-----YDIYGNITAC	1146
Db	1056	KGOQV-----KYVDLTKVNGYT--THVDNDMGNLIVTNKYTPKPKNPIY-----	1101
Qy	1147	HTTFADGTEDHATKFAFNPDPDQTEVHHTHPDMPDNIRLKYDKAGRVINITHNGNTE	1206
Db	1102	-----PEKPKDTP-----PTKPDHNSNKVKZPTPPDXS-----KVDKDDQF--	1145
Qy	1207	N 1207	
Db	1146	N 1146	

ID	YD96	METJA	STANDARD;	PRT;	2894	AA.
AC	Q58791;					
DC	15-MAR-2004	(Rel. 43, Created)				
DT	15-MAR-2004	(Rel. 43, Last sequence update)				
DT	15-MAR-2004	(Rel. 43, Last annotation update)				
DE	Hypothetical protein MJ1396.					
GN	MJ1396.					
OS	Methanococcus jannaschii.					
OC	Archaea; Euryarchaeota; Methanococci; Methanococcales;					
OC	Methanocaldococcaceae; Methanocaldococcus.					
OX	NCBI_TaxID=2190;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=JAL-1 / DSM 2661 / ATCC 43067;					
RX	MEDLINE=96337999; PubMed=8688087;					
RA	Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,					
RA	Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,					
RA	Kerlavage A.R., Dougherty B.A., Tomb J.-P., Adams M.D., Reich C.I.,					
RA	Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,					
RA	Scott J.L., Geoghegan N.S.M., Weidman J.F., Sadow P.W., Nguyen D.,					
RA	Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,					
RA	Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,					
RA	Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;					
RT	"Complete genome sequence of the methanogenic archaeon, Methanococcus					
RT	jannaschii."					
RL	Science 273:1058-1073(1996).					
CC	-!- SIMILARITY: Contains 20 PBH1 repeats.					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/					
CC	or send an email to license@isb-sib.ch).					
DR	EMBL; U67579; AAB99406.1; -					
DR	PIR; C64474; C64474.					
DR	TIGR; MJ1396;					
DR	InterPro; IPR008985; ConA like lec_gl.					
DR	InterPro; IPR001220; Lectin_legB.					
DR	InterPro; IPR006626; PBH1.					
DR	Pfam; PF00139; lectin legB; 1.					
DR	SMART; SM00710; PBH1; 20.					
KW	Hydrothermal protein; Transmembrane; Repeat; Complete proteome.					
FT	TRANSMEM 8 28 POTENTIAL.					
FT	REPEAT 543 567 PBH1 1.					
FT	REPEAT 2085 2107 PBH1 2.					
FT	REPEAT 2135 2156 PBH1 3.					
FT	REPEAT 2158 2180 PBH1 4.					
FT	REPEAT 2201 2223 PBH1 5.					
FT	REPEAT 2224 2244 PBH1 6.					
FT	REPEAT 2245 2266 PBH1 7.					
FT	REPEAT 2267 2289 PBH1 8.					
FT	REPEAT 2290 2311 PBH1 9.					
FT	REPEAT 2341 2363 PBH1 10.					
FT	REPEAT 2367 2389 PBH1 11.					
FT	REPEAT 2390 2419 PBH1 12.					
FT	REPEAT 2422 2444 PBH1 13.					
FT	REPEAT 2455 2477 PBH1 14.					
FT	REPEAT 2479 2501 PBH1 15.					
FT	REPEAT 2512 2542 PBH1 16.					
FT	REPEAT 2550 2582 PBH1 17.					
FT	REPEAT 2589 2611 PBH1 18.					
FT	REPEAT 2612 2633 PBH1 19.					
FT	REPEAT 2638 2660 PBH1 20.					
SQ	SEQUENCE 2894 AA; 322361 MW; 8B64F786537A16DF CRC64;					

Query Match 2.4%; Score 216.5; DB 1; Length 2894;
Best Local Similarity 19.0%; Pred. No. 0.0027;
Matches 383; Conservative 250; Mismatches 719; Indels 659; Gaps 105;
65 LSYSPLNKTIDIGFGIGFGLSVYDRKNSLSLSTGKNTYKVIETDKTVKLOQKLDNLP 124

Db 409 ISYDFVKEII-----IKLESEKIEKLLKRYEKKKYKI-----SKIVKIYNSVIBIKL 460
QY 125 EKD-----LIKE-----NCYRIIHKSGDIEVLTFNNNA-----FDLKV-----KKLL 162
Db 461 KADKEILLKYNLPENILNTTIVKISNKIRVEVNNKVDGWRFSCKIPKGRVKEIV 520
QY 163 NPAGHAI-----YIDNPEATQPLNRIV-----DLDGDHIDPLL-----197
Db 521 GDDGRVIRNNISINRLTEGIVCEVRWVIEN-----NTLYFYDDPIYGYDISLIPAPNHS 575
QY 198 -----NLEVQGLIKTILTFPGQKEGYTELRFN-----RQLNSIHNFSLGNE 241
Db 576 IAVELSYNGQYGGCGQISALVFPYKEDDETTVATYDHAGRTGDYVANNIDAIA-GSK 634
QY 242 NPLTWSFG-----YTPIGKNGILQW-----ITSMTAPGGLKETVNS-----279
Db 635 IAIKYTSGALRQYGVLTAGSLGWTVYVYLSEINRDTDIPLNTVPNGILESVIITDMYAP 694
QY 280 -NNQGHFFPOSANLPVLYTLKQVPGACQALQAEY-----SYTSHNYVG-----327
Db 695 WNNNELN-----ITQKVIIRGNKWFATYIKNPTKTYTLNLFQGMWDMN 741
QY 328 -----GSGNIGWNNKLDNLYGLMTEYNGSTESRRYKDKHGHOIVRIERTYNNYHLLTS 381
Db 742 FRGSWGGDAYNSIDDVWYGDENAPVGDIOYGGFK-----SNIPSEHVDNLYWSTWS 796
QY 382 ECKQON-----GYQTETETA-----YVAI-----IGHNFSQPSQF-----QL 414
Db 797 DIRYDNLNDSYEGDAGTALAWTKSLKPGEIWVVPILWGLGYNTYDMNNEINMGLSQL 856
QY 415 PKT--KTETWESADNSYSEI-----TETTFDESIGN 443
Db 857 YDTGVKSIDYFNNGSGFNPIIPIIYINSITIALYGLVDAYNLVNSINITQINGTYITNS 916
QY 444 PLTKVIKDKTKQKIIISPTHWEYYPAGEVDNCPBPYGFTRFKYKIIQT--PYDSEFKD 501
Db 917 TLINLSVPYEBEKLVS-----FPV--NISNMEPYGAYNIT-----IKTNLPNDQNTSN 961
QY 502 DPERFIQY--RYSILIGSOSHVTLKIEERHYSATOLLNSTLFOYNTDKSELGR-LKOTEC 558
Db 962 DEKILIIYISFSVOPVQEKTNVGEIIF-----YNTILYFVG-----GRDINITYL 1012
QY 559 TKGNGKTYIS-----VVHKEFTYTKQDDTLQOQSHSIITHDNFTIHSQVR 602
Db 1013 TKGWTKIYNNISILIAEDANGDGDWVDINP-NYDLNNSNLPDIYVPTGEINLTVSKTIPS 1071
QY 603 SRYTGRLEFSDTDKDIQTOMSYDKLGRLLTRLTNSGTPYANTLTIDYELN-----NLQDDN 658
Db 1072 TAPLGEI--DTTILKFVNNINPSIFGK--TTFOTSTPYPPSVQKTFYLGHDTLRLTNTS 1126
QY 659 RPPFVITTTDVGNGQL-----RNEFDGAGRHSQCCLKSDGDKGFYTIHTQQYDEQ 709
Db 1127 IPTTINNYTINSLSASWIOYPRFADNFTVVGK-IPILLYINDPNVIFGT-----1176
QY 710 GRHTSTYSVLYTNGROQ-----TDPDKVHLSMS-KSYDNNGQIANTHWSYGVSEKITTVP 764
Db 1177 -EMEKIVVSLMATNGIDSFITGSDVEYLYLDDTIKSY-----IPN-----ITLDS 1220
QY 765 I-----TLTATKLOQSNVNVQKGEVTTVTPSCQPIQITLFPDEAGHLOSCHT--L 813
Db 1221 IITIPKNYVLRVENQIISNSINIYEN---STY-PSNIYINLTITTYVNVNIFSDKNVYL 1276
QY 814 TRDGWDRVKETDAIGCTIYQYDYNRVITQITLPDGTI-VNR-----KVAP-----859
Db 1277 PNEVNTIFANITDPIG-----SYDISGANITVYVYVNGSVVINSMLLQEIKNKSPSLWL 1331
QY 860 -----FSTD-----TLTIDIRVNGI-----SLGQOFDGLSRLTQSDGG 894
Db 1332 YNVSFSLPESKYLITITIGIESNGVISKKNYTYCGVEIOGYVKEDFGTLGK-EDSEDKG 1390
QY 895 RVWAYTYS-----AGNDQCP---STVI--TPDQGFHYQ---YQPELDDAVL-----933

RESULT 8

P2P LACLC

ID P2P LACLC

AC P15293

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DE 28-FEB-2003 (Rel. 41, Last annotation update)

DE PII-type proteinase precursor (EC 3.4.21.96) (Lactocypin) (Cell wall-

DE associated serine proteinase) (LP151).

GN PRT.

Db 1391 -IYGVNVSLLENNNDGIPDIDGTIVNSTTTDFGHYSFELVYNSKTYFVVVNSRTVGT 1449
QY 934 -----OVASNEI-----TOQFSYNPVTGALLKVAEGOS-----LTPIYVPSGRL-- 973
Db 1450 RGLNPOYSKNDIWAEBTYQTVTPINSS--QWIANGNASIFPDKLLTTDDTGEAGSVWY 1507
QY 974 -KXENINDMKMSYMLWTLRGLNGYDLDLTGTQIKISRD-----HGRVTQIKDSSI 1023
Db 1508 YKPVNLSEDLVVEFYAYLGDNDPDGADGITFTLQSLGTNELGGTGLGYGGISPSVAVEV 1567
QY 1024 KTTLN-YD-----DLNRHIGSQVTDLATGCHMLTTTVEFDGLNREIGRKLCDSSGH 1072
Db 1568 DTWLNDFDAPATDTHAIDVNGINHTYNSLT-----YPTPNYDIDLVNEDGRE-----H 1617
QY 1073 TLDIOQSWLKTOQLANRIVKLVNGVLTQTEQSYDSNRNLNOYKCDGAECTDKYGHISVT 1132
Db 1618 LIKI--VWVAT-----TKLQVYFDGNLSLTWNK-----DIT 1647
QY 1133 ONFTYDIYCNITACHTTTADGTEDHATFKFANPT-----DPCLTEVHHT-----HPD 1180
Db 1648 Q-----ILGN--SAYPGFTGGTGGAQNUQYVVKPIVYKNGDGYIINPTYGVMFEGORDN 1700
QY 1181 MPDNIR-LKYDKAGRVINITDNEGN--TENFTYDILGRLOQ-GQGS-----VYGY 1226
Db 1701 EEDNWDGKYEYH-CLINLNSYSYGNITFGPSFDVITNTKSTGQGSFSQFIKNANAIYK 1759
QY 1227 DPLNRLVSQKTDLDCELYRETMLVNEVRNGEMIRLLRTGETII-----AQRASK 1278
Db 1760 D-----ESYFRIPNI--DAKNGHY-IVTSGNKILDLNLTIVNGSTQINGT 1801
QY 1279 VLLTG-----TDSQOSVILTSDKQNLQSEAYSAVGKHKSTANDASI-----LG 1321
Db 1802 IILSGLOWTANGVAYNNNSNLTILTPDDYN--QKGSVWYKYPVNLSEDLVVEFYAYLG 1859
QY 1322 YNGERADPVS-----GVTHLGN-----GYRSYDPT-----LMRHTP--DLSLSP 1359
Db 1860 DNPDCADGITFTLQSLGTNELGCTGDLGYGGSVAVEVDTLNDFDPSATTDHAIAD 1919
QY 1360 GAGGIN-----PYSYCLGDPINRSDPSGHLSSQAWTGTGICWGIAGLLLTATGGWA 1409
Db 1920 VQGNLNTYNSLTYSPTNPYDLG--NVEDGREHLIKIVN-----1957
QY 1410 IAAAGGIAAAIATSTTALAFG---ALSVTSDITSIVSGALEDASPKASSILGWVSMGMG 1466
Db 1958 -----ATTKTQVYFDGNLALTWNKDITQIGN-----STYFGTGGTGG 1997
QY 1467 AAGL--ABSAIKGCTKLATHLGAFAEDGENALLKSTSESSIKWGTVRSLDREIVRNEE 1523
Db 1998 AKNLQVVKPYVYKNGDNLN-----LEEISPNPIDNVGADTYIGNIFFENVSUGILGNET 2053
QY 1524 G-----QVIKDHSG-----YTDNFM--GKGEQAILVHGDKDGF 1556
Db 2054 GLNNLTLSKSGIYVKILNAGVKLVYDWSLQNYPIYIDLNTINASGGYGISMLNKIWL 2113
QY 1557 YHTEGNKNGKQPY-----TRHTPEQLVDYLDKNNIYDLTQGGDKPVHL 1600
Db 2114 YNSQISLKNVGVIYWANWAGFGNITTYNITSSNOGLVLYKDGNGIKLINSOIKNSVY 2173
QY 1601 LSCYKGSKAADKMAKYINRPVIAYSNKPIT 1631
Db 2174 EGYSKNSTLEILNSIINNSIGIYANISSI 2204

Db 1304 NSVTFDQGVTFGEAFNATSAKFDPKGTATITGKVKHPTTLQVDGKQIPI-----KDD 1359
Qy 1131 VTQFTYD-----IYGNITACHT-----TF-----ADGTEDHATKRFAN 1164
Db 1360 LTFSTLTLGLGQKPGVGVGDTTQNTKTFQEAITFILDVAFTLSLDSSTDPVIT-NW 1418
Qy 1165 P-----TDPQCLTEVH-----HTHPDMPDNIIRLYKAGRV 1195
Db 1419 PNQITGTATDQAQYLSLSINGSSVQSYVDININSKPGHMAIDQP--VKLLEGKNVLT 1476
Qy 1196 INITDNEGN--TENFTYDTLGLQNGQSVYGYDPLNRLYSQKTDTLDCELYRETNMLVN 1253
Db 1477 VAVTDSEDNTTKIIT-----VI-YEPKTKLAAPTVPSTTTPAKTVITLTA 1522
Qy 1254 EVRNGEMIR-----LLRTGETIIA-----QQRASKVLLTGTDSOQ 1288
Db 1523 SAATGETVQVSADGGKTYQDVPAAGVTVTANGTFKFKSTDLYGNEPSPVYVVTNIKADD 1582
Qy 1289 SVILTSOKNL-----SQEAYSACKH-----KSTAND---ASILGNGER 1326
Db 1583 PAQLQAKQELNLIASAKTUSAGKYDDATTALAAATQAKQATDQTNASVDSUTGAN 1642
Qy 1327 AD-----PVSQVTHLNGYRSYDPTLMRFHTPDSLSPFGAGGINPYSCIGLDP 1374
Db 1643 RDLQTAINQLAAXLPADKKTSLNLQLOSVKAALETDLGNQTDSTG-----KTFEALDDL 1698
Qy 1375 INRSDPGLHSQWAWTGIGMGIAGLLTIATG-----GMAIAAGGIA--AIAAST 1423
Db 1699 VAQAQAGTQDDQQLAKVLDLAVLAKLAEGINKAATPAEVGNAKDAATKTYADIAT 1758
Qy 1424 STTALA-----FGALSVTSDITSIVSGALEDASPKASSILGWVGMGAAGLAESAIAK 1476
Db 1759 LTSQASADASDKLAHLQAQSLKTKVAAAVEAAKT-----VGKGDSTTTSKGGG 1810
Qy 1477 GGTKLATHGAFADGNALLKSTSSSRIRKGVTRSLREIVRNBEGQVHKDHSRGYTD 1536
Db 1811 QGTPAPTPGDIQKDKGEGSQPSGGNIPNPAITTSSTDDTDRNGQLTS-----1862
Qy 1537 NFMKGK 1542
Db 1863 ---GKG 1865

RESULT 9

P3P_LACLC STANDARD; PRT; 1902 AA.
ID AC P15292; 1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE P111-type proteinase precursor (EC 3.4.21.96) (Lactocepin) (Cell wall-
GN associated serine proteinase).
GN PRTP.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RC STRAIN=SK11.
RC MEDLINE=89340435; PubMed=2760036;
RX Vos P., Simons G., Slezien R.J., de Vos W.M.;
RT "Primary structure and organization of the gene for a procaryotic,
RL cell envelope-located serine proteinase.";
RL J. Biol. Chem. 264:13579-13585(1989).
CC -I- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -I- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC specificity, although some substrate preference have been noted,
CC e.g. large hydrophobic residues in the P1 and P4 positions, and
CC pro in the P2 position. Best known for its action on caseins,
CC although it has been shown to hydrolyze hemoglobin and oxidized

CC insulin B-chain.
CC -I- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (Potential).
CC -I- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J04962; AAA03533.1; ALT_SEQ.
DR HSP; P00782; 2SST.
DR MEROPS; S08.019; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR002029; Peptidase_S8.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00336; SUBTILASE_ASP; 1.
DR PROSITE; PS00337; SUBTILASE_HIS; 1.
DR PROSITE; PS00338; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Zymogen;
KW Signal; Plasmid.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1870
FT PROPEP 1871 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT SITE 1867 1871
FT MOD_RES 1870 1870
FT SEQUENCE 1902 AA; 200550 MW; 87CEBAA9345F9D3 CRC64;
SQ
Query Match 2.3%; Score 205.5; DB 1; Length 1902;
Best Local Similarity 18.7%; Pred. No. 0.0055;
Matches 316; Conservative 195; Mismatches 589; Indels 586; Gaps 83;

Qy 239 GNENPLT---WSFGYTPIGKNGILQ---WITSATPG-GLKEIV-----NYSNNQ- 284
Db 521 GTATPMTSIALTTTPTTGLSVTGQKLVDWVTAHPDSDSLGKVTLAMLPNQKYTEDKMS 580
Qy 285 H--HFPQSANLPVLPVTLMKQVPGAGQPAIAEYSYTSNHYVG-----GSGNGIWNK 336
Db 581 DFTSYGVSNSLSFKPDIT---APGNTWSTQNNNGYTNMSGTSMASPIAGSQALLKQA 636
Qy 337 LDNLVGLMTEYNGSTESRRYKDEG---HQIVRIER-----TNNYHLLTSECK 384
Db 637 LNNKNPFYAY-----YQKLGATLDTFLKVTENMTAQPINNINN--VIVSPRR 685
Qy 385 QQNGYIQ-----TTETAYVAILIIGHNFDSPQSFQFOLPKTKTETWRSADNSY 429
Db 686 QGAGLVVDKAAIDALEKNPSTVVAENGYPVELKDFSTDKTKLTFNTRTHELTYQMD 745
Qy 430 RSEITETTFDSGNPLTKVKIKDKTKIISPTSHWEYVFPAGEVNDCEPE-----PYG 482
Db 746 SNTDTNAVYTSATDPNSGLYDKK-----IDGAAIKAGSNITVPAG 786
Qy 483 FTRFVKKLIQTPYDSEFKDDSEKFIQRYSLIGSQSHVTLAIEERHYSATOLLNSTLPQY 542
Db 787 KTAQTEFTLSLPKSPDQOQFVEGLNF-----813
Qy 543 NTDKSELGRLLKQTECKGKNGKTYSVVHKFTYKQDDTLQOQSHITTHDNFTIHRSQVR 602
Db 814 -----KSGDG-----SRLN 822

QY 603 SRYTRFLPSDDTDKDIYVQM---SYDKLGR-----LLTRTLNSGTPYANTLTIDYELNLL 654
Db 823 LPYMG-FFGWMDGKIVDSLNGITYSPAGNFGFVPLLNKNTGTQYGGMVTADGNKT 881
QY 655 QDDNRPPIVITTDVNG-NOLRNEF-----DGAGRHSVQCLKSDSGDGKF 698
Db 882 VDDQAIAP---SSDKNALYDISMKYLLRNI SNVQVDILDGQGNKVTLLSSSTNEKKTY 938
QY 699 YTIHTQ-----YDEQGRHMTYSO-----YLTNG-----RQOTD----- 729
Db 939 YNAHSQYIYNAPADGTYYDQDNIGKTADDGSYTYRISGVPEGGDKRQVDFVPFKLD 998
QY 730 ---PKVHLSKSYDNWQIANTHWSYGVSEKITVDPIITLTATKLOLSNNVQTKEV 786
Db 999 SKAPTVRHVALSAKTEN-GKT-----QYLTAEAKDLSGLDATKSVKTEINEV-TNLDA 1051
QY 787 TTYTPSQOPIQITLFEAGHLQSCHTLTRDGDWRV-----KETDAIG-----QCTIYQYD 837
Db 1052 T-----FTDAG-----TTADGYTKIETPLSDEQAQALNGDNSAELYLTD 1091
QY 838 NYNRVIQITLPDGTIVNRKYPAPSTDITLITRVNGISLGCQTFDGLSRLT-----QSOD 892
Db 1092 NASNA---TDQASV---QKPGSTS---FDLIWGGGIP-----DKISSTTGYEANTIQ 1137
QY 893 GGRVMAVTSYAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEITQOFSYNPVTA 952
Db 1138 GG---TYTFSG---TYPAAV---DGTVDAGKKHDLNTTYDAATNSFTASM---PVTNA 1185
QY 953 LLKAVAEQSLTPYIPSGRLKMNINDMKMSVLTWLRGLENGYDITLGTQIKISRDTH 1012
Db 1186 DYAAQVD-----LYADKAHTQLLKHFDTK-----VRLMAPTFDL-----KFNNGS- 1226
QY 1013 GRVTOIKDSSIKT---TLNYDNLNHHIGSQVTDLATGMLTTFVDFGLNREIRKGLCDSS 1070
Db 1227 ---DOTSEATIKVGTVSADTKTVNGHTVAALDAQHHFSVDPVNVYGDNTIKVATDKD 1283
QY 1071 GHTLDIQO---SWLTKQOLANRIVKLVNGVLTQRTQYS-----YDSNRRL----- 1111
Db 1284 GNTTEQKTISSYDPPDKLSVTFDQGVKFTGNFNKFNATSAKFPDPTGATITGKVXHP 1343
QY 1112 -NQVKDCACBPTKYGHSIVTQFTYDI-----YGNL-----TACHTTADGTE----- 1155
Db 1344 TTTLOVDGKQPI-----KDDLTFTFETLDLGLGKPGVVVGDITQNTQKTEALSFLDA 1399
QY 1156 -----DHATFKFANPTDP-----CQLEVEH-----H 1176
Db 1400 VAPTLSDSSTDAPVYTNDRPNFQITGATDINAQVLSLSINGSSVASQVEDINNSGRPGH 1459
QY 1177 THPDVFNRLKYDKAGRVINITDNHGN---TENETDYLGLRQNGSVYGYDPLNRLVS 1234
Db 1460 MAIDQP---VKLEGGKVLTVAVTDSEDNTTKNIT-----VY-YEPKKTAA 1503
QY 1235 QKTDLTDCELXYRETMVNEVNGEMIR-----LLRTGETIIA----- 1272
Db 1504 PTVTSTTEPAQVTLTANAATGETVQYSADGKTYQDVPAAGVITANGTFKFXSTDL 1563
QY 1273 ---QORASKVLLTGTDSQSVILSDKNL-----SQEAYSAYGH----- 1310
Db 1564 YGNESPAVDYVYVNTIKADDPQAQQAQOELTNLIASAKTUSASGKYDDATTALAAATQK 1623
QY 1311 KSTAND---ASILGVNGERAD-----PVSQVTHLNGYRSDYDPTPLMFHTPDS 1355
Db 1624 ACTALDQTNASVDSLTGANRDLQTAINQLAALPADKXTLSLNQLQS-----VKDA 1674
QY 1356 LSPFGAGGINPYSYCLGDPINRSDPSGHLSSQAWTGICNGAGLILLIATGMAIAAAGG 1415
Db 1675 LGT-----DLGNQTDPS-----TGKTFTAALDULVAQQAQ 1705
QY 1416 -----IAAAIATSTTALAFALGSVTSDDITSIVSGALEDASPKA---SSILGWYSMGMA 1467
Db 1706 TOTDQLOQATLAKILDEVLAKLAEGIKAAATPAEVCNKAADATGKTWADIADTLTSGAS 1765
QY 1468 AGLAESAIKGGTKLATHIGAFAEDGENALLKSTSESSRIKGVTRSLDRIVRNEEQVI 1527

Db 1766 ADASD-----KLA-HLQAL-----QSLKTKVAAAEAAKTGVKG----- 1798
QY 1528 KDHSGYTDNFMGKEQAILVHGD---KDGFLYHTEGNKHNGK-----PYTRHTPEQLVD 1580
Db 1799 -DGTITGTDKGGGGGTGA-PAPGDTGKD---KSGDEGQSPSGGNIPTKPAF---TTSSTTD 1851
QY 1581 YLKNMNIVDLTQGGDKPVLHLLSCYKSGSAAADKMAKVINRVIAYSNKPTIS---OCLA 1636
Db 1852 DTTDRN-QOLTSQ-----XGALPKTGETTERPAFGFLGVIVVSLMGLGLK 1896
QY 1637 RIERKD 1642
Db 1897 RKOREE 1902
RESULT 10
PIP LACLC
ID PIP LACLC STANDARD; PRT; 1902 AA.
AC P16271;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE P1-type proteinase precursor (EC 3.4.21.-) (Wall-associated serine
DE proteinase).
GN PRTP.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WG2;
RX MEDLINE=88149035; PubMed=3278687;
RA Kok J., Leenhouts K.J., Haandrikman A.J., Ledeboer A.M., Venema G.;
RT "Nucleotide sequence of the cell wall proteinase gene of
RT Streptococcus cremoris Wg2";
RL Appl. Environ. Microbiol. 54:231-238 (1988).
CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC -!- GROWTH OF THE BACTERIA ON MILK.
CC -!- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC e.g. large hydrophobic residues in the P1 and P4 positions, and
CC pro in the P2 position. Best known for its action on caseins,
CC although it has been shown to hydrolyze hemoglobin and oxidized
CC insulin B-chain.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M24767; AAA17677.1; --
CC HSSP; P00782; LS01.
DR MEROPS; S08.019; --
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00882; Peptidase_S8; 1.
DR PRINTS; PRO0723; SUBTILISIN.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.

KW Hydrolyase; Serine protease; Cell wall; Peptidoglycan-anchor; Zymogen.
KW Signal; Plasmid.
FT PROPEP 33
FT CHAIN 187
FT PROPEP 187
FT ACT SITE 217
FT ACT SITE 281
FT ACT SITE 620
FT ACT SITE 1867
FT MOD RES 1870
SQ SEQUENCE 1902 AA; 199910 MW; 2901C7F19B2E5D0B CRC64;

Query Match 2.3%; Score 201; DB 1; Length 1902;
Best Local Similarity 18.6%; Pred. No. 0.0097;
Matches 311; Conservative 195; Mismatches 600; Indels 562; Gaps 79;

QY 205 IKTLTLPQGEYRTELFLNQLNSIHFNLSGNENPLTWSFGYTPICKNGILGOWIT 264
DB 562 VKIALTLVPNQY-----TEDKMSDTSYGVSNL-----SFKP 595
QY 265 SMTAPGLKETVYNNNOGHFFQSANLPVLPVTLMKVPGAGQPAIOAEYSYTSYNY 324
DB 596 DITAPGG---NIWSTQNNNG-YTNMSGTSMASPPF-----AGSQALLKQ----- 635
QY 325 VGGSGNGIWNKLDNLGLATEYNYGSTEERRYKKEG---HQIVRIER-----T 372
DB 636 -----ALNNKNPFYAY-----YKQKGTALTDLTKTVMNTAQPINDIN 675
QY 373 YNNYHLTSECKQONGYIQ-----TTETAYVAIIGHNFDSPQSQFOLPKT 417
DB 676 YNN--VIVSPRQAGLVVDKAAIDALEKNPSTVVAENGYPAVELKDFSTDKTKLTFT 733
QY 418 KTEWRADNSYREIETTFDESGNPLTKVYKDKTKQKILSPSTHWEYPPAGEVDNCP 477
DB 734 NSTTHELTQMSNTDNAVYTSATDENSGLYDKK-----IDGAA 774
QY 478 PE-----PYGFRFRVKKIQTYPDEFKDDPKFYQRYSLIGSQSHVTLKIEERHVS 530
DB 775 IKAGSNITVAGKTAQIEFTLSLPKSPDQOQVEGFLNF----- 813
QY 531 ATQLNLTLPYNTDKSELGRLLKQTECKGNGKTVSVVHKFTYTKQDDTLQOQSHSITT 590
DB 814 -----KGSQD----- 818
QY 591 HDNETHRSQVRSRYTGLPSDDTDKDIVTQM-----SYDKLG-----RLLTRINSQTPY 641
DB 819 -----SRNLPEYMG-FFGDMNDGKIVDSLNGITVSPAGNGFTVPLLTNK-NTGTQY 868
QY 642 ANLTLYDELNNLQDDNRPPFVIITTDVNG--NQLRNEF-----DGAGRHY 685
DB 869 YGGMVTDADGNQTVDDQAIAP---SSDKNALNDISMKYVLLANISNVQVDILDGQGNKY 925
QY 686 SQCKDSGDGKFTYIHTQO-----YDEQRHHTSYSD-----YLTNG----- 724
DB 926 TTLSSTNLTKTYTNAHSQQYIYNAPAWDGTYYDQSDGNIKTADDSGYTYRISGVPEGG 985
QY 725 --RQOTD-----PKVHLSKSYDNNWQGTANTHWSYGVSEKITVPITLTATKOL 773
DB 986 DKRQVFDVPPKLDKSAFTVRHVALSAKTEN-GKT-----QYLTAERAKDLSGLDATKSV 1039
QY 774 QSNENNVQTKGVETTPSPQOPIQITLFDAGHLQSCHTLTRDGRWRV-----KXTDAI 828
DB 1040 KTAINEV--TNLDAT-----FTDAG-----TTADGYTKIETFLSDEQAQAL 1078
QY 829 G-----QCTIYQYDYNRVQITLPDGTIVNRKYAPFSTDLITDIRVNGISLQQTDFGL 884
DB 1079 GNGNSAEYLTDNASNA--TDQDASV---QKPGSTS---FDLIVNGGGIP-----DKI 1124
QY 885 SRLT-----OSQDGRWYATYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLOVASNE 939
DB 1125 SSTTTGYEANTQGGG---TVTFSG---TYDAAV---DGTVTNAQGGKHLNTTYDAATNS 1175

QY 940 ITQFSYNPVTGALLKAVAEQSLTPIIYPSGRKLKQENINDMKMSYLWTIRGLENGYTD 999
DB 1176 FTASM---PVTNADYAAQVD-----LYADKAHTQLLKHFTDK-----VLTAPTFTD 1219
QY 1000 LGTTTQKISRDRTHGRVTAQIKOSIKT--TLNYDDLNRHIGSOVTDLATGCHMLTITVEDG 1057
DB 1220 L-----KFNNGS-----DQISEATIKVTGVSADTIVNVGDTVAALDAQHHSFVDPVNY 1270
QY 1058 LNREICRKLCDSSGHTLIDQ---SWLKTQQLANRIVKLVGLQTEQYS-----YDSR 1108
DB 1271 GNTIKVATDEGNTTTEQKITISSYDPDMLKNSVTFDQGVTFGANEFNATSAKFYDPK 1330
QY 1109 NPL-----NOYKCDGAECPDKVYHSIVTQNTYDI-----YGN-----TACH 1147
DB 1331 TGIATITGVKHPTTTLQVDGKQIPI-----KDDLTFSTLDTLGLTQKPFVGVGDTTQN 1386
QY 1148 TTFADG-----TEDHATEKFNPTDPCOLT-----EVHHTHPMDNIRLKYD---KAGR 1194
DB 1387 KTFQEAFTILDVAPTLSLESSTDAVYTNDPNFQITGATDNAQYLSLSINGSSVASQ 1446
QY 1195 VINITDNHNTENTFYDILGRLOQCG-----SVGYDPLNRLVSQK 1236
DB 1447 YVDIINSGKPGHMAIDQPVKLEGNVLTVAVTDSEDNNTTKNITVY-YBPKTKLAAPT 1505
QY 1237 TDTLCELYYRETMVNEVRGEMIR-----LLRTGETIIA----- 1272
DB 1506 VTPSTTEPAKTVTLTANSAATGETVQYSADGGKTVQDVPAGVTVTANGTFPKXTDLYG 1565
QY 1273 -QORASKVLLTGTDSQQSVILTSDKQNL-----SQEAYSAYGKH-----KS 1312
DB 1566 NESPAVDYVVTNIKADDPQAQQAQKQELTNLIASAKTILSASGKYDDATTTTALAATAQKAQ 1625
QY 1313 TAND---ASILNGYGERADPYSGVTHLGNYSYDPTLMRPHTPDLSLSPFFGAGGINPYSY 1369
DB 1626 TALDOTNASVDSLTGANRDLQATINOLAAPADPKKTSLLNLOLQSVKAAALGT----- 1677
QY 1370 CLUGDIPNRSDDPSGHLSSWAQWTGIGMGIAGLLLTIATGMAIAAAGGI-----AAAAIAT 1423
DB 1678 ---DLGNQTDPS-----TGKFTALDDLVAQQAQAGTQTDQHQATLAKV 1719
QY 1424 STTALAFGALSVTSDITSIVSGALEDASPKA---SSILGWVSMGMAAGLAESAIGKGTKL 1481
DB 1720 LDAVLAKLAEGIKAAATPAEYGNNAKDAATGKTWYADIADTLTSGQASADSD-----KL 1772
QY 1482 AYLHGAFAEDGENALLKSTSESSRIKMGVTRSLDREIVRNEBQGVIKDHSRGYTDNFMWK 1541
DB 1773 A-HLOAL-----QSLKTKVAAAEAAKTVGK-----DGTGTSKGGGQ 1811
QY 1542 GSOAILVHGD--KOGFLYHTEGNKNGK-----PYTRHTPEQLVDYILKDNINVDLTQGG 1594
DB 1812 GTPA-PAPGDIKGD---KGDEGQSPSGGNIPTNPAT--TTSTSTDDTTDRN-GQLTSG- 1863
QY 1595 DKPVHLLSCYKSGSQAADKMAKVINRPVIAYSNK-PTISQGLARIER 1641
DB 1864 -----KGALPKTGETTERPAFGFLGVIVILMGVLGKRK 1898

RESULT 11
P2P_LACPA STANDARD; PRT; 1902 AA.
ID_P2P_LACPA Q02470;
AC Q02470;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE PII-type proteinase precursor (EC 3.4.21.96) (Lactocepin) (Cell wall-
associated serine proteinase) (LPI51).
GN PRTP.
OS Lactobacillus paracasei.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1597;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=NCDO 151;
 RX MEDLINE=92381481; PubMed=1512565;
 RA Holick A., Naes H.;
 RT "Cloning, sequencing and expression of the gene encoding the cell-
 RT envelope-associated proteinase from *Lactobacillus paracasei* subsp.
 RT paracasei NCDO 151";
 RL J. Gen. Microbiol. 138:1353-1364 (1992).
 RN [2]
 RP SEQUENCE OF 189-196.
 RX MEDLINE=92226694; PubMed=1564442;
 RA Naes H., Nissen-Meyer J.;
 RT "Purification and N-terminal amino acid sequence determination of the
 RT cell-wall-bound proteinase from *Lactobacillus paracasei* subsp.
 RT paracasei";
 RL J. Gen. Microbiol. 138:313-318 (1992).
 CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
 CC GROWTH OF THE BACTERIA ON MILK.
 CC -!- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
 CC specificity, although some substrate preference have been noted,
 CC e.g. large hydrophobic residues in the P1 and P4 positions, and
 CC Pro in the P2 position. Best known for its action on caseins,
 CC although it has been shown to hydrolyze hemoglobin and oxidized
 CC insulin B-chain.
 CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 CC an amide bond (Potential).
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M83946; AAA25248.1; --
 DR PIR; B44858; B44858.
 DR HSSP; P00782; 1801.
 DR MEROPS; S08.019; --
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR00137; PA.
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR Pfam; PF02225; PA; 1.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR TIGRFAMs; TIGR01167; LPTXG anchor; 1.
 DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Zymogen;
 KW Signal.
 FT SIGNAL. 1 33 POTENTIAL.
 FT PROPEP 34 187
 FT CHAIN 188 1870
 FT PROPEP 1871 1902
 FT ACT_SITE 217 217
 FT ACT_SITE 281 281
 FT ACT_SITE 620 620
 FT ACT_SITE 1867 1871
 FT MOD_RES 1870 1870
 SQ SEQUENCE 1902 AA; 200253 MW; D8C9F38CE5DA582 CRC64;
 Query Match 2.3%; Score 200.5; DB 1; Length 1902;
 Best Local Similarity 19.2%; Pred. No. 0.01; Mismatches 659; Indels 645; Gaps 94;
 Matches 361; Conservative 215;
 12 ITMSDNNNEFFQANNFTSAVSGVDPRPTGLYNIQITLGHVIGNGNLGPTLPPLTSLSPN 71
 419 VASAENTDIVISQAVTIID-----GKDLQIGPETIQLSSNDFTGS----- 457
 72 KTDIGFIGNFGLSVYDRKN-SLLSLSTGENYKVIETDKTVLQOK-----KLDNLRPEK 126

Db 458 -----FDQKFYVVKDAGSLSKGAAADYTADAKGKIAIVRGELNFAF 501
 QY 127 DLKENCYR-----IIHSG-----DIEVTGFNNNAFDLKVPKKLNPAAGHAIYI 171
 Db 502 KQK---YAQAAGAGLIIVNNDGTATPLTSLRTLTPTFGLSSTKQGL-----V 549
 QY 172 DWNFEATOPRLNRIYDDLGDHDIPLNLEYOGLIKTLTLPQOKEGYRTELRFNLQNL 231
 Db 550 DW--VTAHP-----DSSLG-----VKIALTLPLNQY----- 574
 QY 232 SIHNFSLGNEPLTWSFGYTPIGKNGILGOWITSMATPGGLKETVYNSNNQGHFFQSA 291
 Db 575 -----TEKMSDFTSYGPVSNL-----SEKPDITAPGG---NIWSTQNNNG-YTNMSG 618
 QY 292 NLPVLPYVTLMKQVPGAGQPAIQAEYSYTSHYVGGSGNGIWNKLDNLXGLMTEYNYGS 351
 Db 619 TSMASFFI-----AGSQALLKQ-----ALNNKNPFFA----- 646
 QY 352 TESRRYKDKEG---HDQIVRIER-----TYNNYHLTSECKQNGYIQ----- 391
 Db 647 ---DYKQKGLTALTDLTKTVMNTAQPINDINYN--VIVSPRQAGLVDVKAADAL 700
 QY 392 -----TTETAYVAIIGHNFDSPSQPOLPKTKETWRSADNSYRSEIETTFDESGNP 444
 Db 701 EKNPSTVVAENGYPFAVELKDFSTDKTKFLTFNRTTHELTQYQMSDNTDINAVTSDTP 760
 QY 445 LTKVIKDKKTO---KIISPSTHWEYPPAGEVDCPEPYGFTFRVKKIIQTPEYSEFK 500
 Db 761 NSGVLYDKIDGAALKAGSDIT-----VPAGKTAQIEFTLSLSPKSDQ 804
 QY 501 DDEPKFIQRYSLIGSQSHVTLKIEERHYSATQLLNSTLFOYNTDKSELGRLKQTECTK 560
 Db 805 QFVEGFLNFKGS-DGSR-----LNLPMGFFGDMND-GKIVDSL----- 841
 QY 561 GENGKTYSVVHKFTYTKODDTLQOSHSLTTHDNTFIHRSQVRSRYTGLRFSDDTKDIVT 620
 Db 842 --NGITYSPAGNGYGV-----PLLTWN-TGH-----QYGGWTDADGKTVD 883
 QY 621 Q---MSYDKLGRLLTRLNSGTPYANLTVD-----YELNNLODNRPPVIITTDVNG 671
 Db 884 DQAIASFSDK-----NALYNDISNQYLLRNISN-----VQVDI-- 917
 QY 672 NQLRNEFDGAGHYSOCLKSDGDKKFTYIHTQ-----YDEQGRHSTYSYD 719
 Db 918 -----LDGQGNKVTTLSSTNTQKTYDAHSQKIYYNAPWDGTYDQDGNIKTADD 971
 QY 720 ---YLING-----ROOTD-----PDKVHLSMSKSYDNWGOIANTHWSYGVSEK 759
 Db 972 GSYTYRISGVEPBGDKRVDFVPFKLDSKAPTVAHVALSAKTEN-GKT-----QYLYTAE 1025
 QY 760 ITVDPIITATKQIOLSNNSNNVTGKEVTTYTFPSQOIQLTDFEAGHLQSCHTITRQWD 819
 Db 1026 AKODLSGLDATKSVKTAINEV-TNLDAT-----FTDAG-----TTADGYT 1064
 QY 820 RVR-----KETDAIG-----OCTIYQYDNNYRVIQITLPDGTIVNRKYAPFSTDLITDIR 870
 Db 1065 KIETPLSDEQAQALGNGNSAEYLITDNASNA---TNQDASV---OKFGSTS---FDLI 1114
 QY 871 VNGISLGQOQTFDGLSRLT-----SQDGRVWATYTSAGNDOCPSTVITPDQGIHQYQ 925
 Db 1115 VNGGIP---DKISSTTTGYEANTQGG---TYTFSG---TYFAAV---DGTITDAGK 1161
 QY 926 PELDDAVLQVASNEITQQFSYNPVTGALLKAVABGQSITPIYPSGRKXENINDMKMS 985
 Db 1162 KHDLTNTTYDAATNSFTASWA---VTNADYAAQVD-----LYADKAHTQLLKHFDTK--- 1209
 QY 986 YLWTLRGLENGVDLTGTIQKISRDTGHRVTVQIKDSSIKT---TLNYDDLNRHIGSOVTDL 1043
 Db 1210 ---VELTAPTFTDL-----KFNNGS-----DOTSEATIKVTGVSSDTKTNVGDTVAAL 1256
 QY 1044 ATGHEMLTITVVEFDGLNRIEGRKLCDSGHTLIDIQOSMLKTCQLANRIVKLINGVLQRTQY 1103

Db 1257 DAQHFSVDVFNVDNTIKVTATDEGNTTTEQKTISSYD-----PDVLKNVTF 1308
QY 1104 SYDSRNELNOVKDGAECPTDKVGHSTVQNTFTYDIYGNITACHTTF-ADGTE---DHA 1158
Db 1309 DQGVKFGANFNATSAKFDPKTGIAIT-----GKVKHPTTTLQVDGKQISIKNDL 1360
QY 1159 TFKFANPTDCQLTEVHHPTDMPDNLRLKYDKAGRVINITDNHGNENTFYD----- 1211
Db 1361 TFSPT-----LDLGLTGCKPFGVVGDT-----TQNTFOEALTFIL 1397
QY 1212 -----TLGLRQNGQSGYGVDPNLRLVSQKTDLDCELYYRETMVNEVNGEMIRLRT 1266
Db 1398 DAVAPTLSDSDTAPVYNDPFOITGTAID-----NAQVLSLAIN 1439
QY 1267 GETIAQGRASKV-----LITGDSQOSVILTSKQNLSEAYSAVGKHK 1311
Db 1440 GSHVASQYADININSGKPGHMAIDQPVKLEGG-KNVLTVAVTDSNNNTTKITVYEPK 1498
QY 1312 S-----TANDASILGNCE-----RAD-----PVSGVTHLNG 1339
Db 1499 KTLAAPTIVTSTTEPAKTVTLTANAAA-----TGTVQYSADGGKTVQDVPAAGVTVTANG 1554
QY 1340 YRSYDPTLMPFHTPDSLSPFGAGINPYSYCLGDPINRSPSG-HLSWQAWTIGMGIAG 1398
Db 1555 -----TFKFKSTDLY-----GNESPAVDVYVNTNIKADDPQAOLQAKQALTNL---IAS 1599
QY 1399 LLLTIATGMAIAAGGIAAIASTSTTALAFGALSVTSDITSIVSGALED----- 1449
Db 1600 AKTLASAGKXDDATTTALAAATOKAQOT-----ALDQTDASVDSLGTANEDLQTAINQLA 1653
QY 1450 ---ASPRASSILGVWSMGMGAG-----LAESAIGKGTGLATHLGAFAEDGENALLKST 1500
Db 1654 AKLPADKKTSLNLOLQSVKALGTDLGNQDTPSGKFTTAAALDDLVAQAAG-----T 1706
QY 1501 SESSRIKGVTRSJDRBIVRNEEG-----QV---IKHRSRG---YTD---NFMKGGEQAI 1546
Db 1707 QTADQLQASLAKVLDAVLAKLAEGIKAAATPAEVGNAKDAATGKTWYADIADTLTSGQAS- 1765
QY 1547 LVHGDCKGFLVH-----TEGKNKNGKGFYTRHTPEO-----LVDYL 1582
Db 1766 ---ADASDKLAHLQALQSLKTKVAVAEAAKTAGKDDTTGTSKGGQGTAPAPGDTG 1822
QY 1583 KD--NNIVDLTQGGD---KPVHLLSC-----YKSSGAADKQAKYINRPVIAYS 1626
Db 1823 KDKGDEGSPSSGGNIPKPAITTTSTDDTTDSNGQHTSGKGAIPKTAETTERPAPGFL 1882
QY 1627 NKPTIS-----QGLARIERKD 1642
Db 1883 GVIVYSLMGLGLKRRQREE 1902

RESULT 12

ID YEEJ_ECO57 STANDARD; PRT: 2660 AA.
AC Q8X8V7; Q8X2B9; Q8X2C0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yeeJ.
GN Z3135 OR_ECS2775/ECS2776.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11205551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,

RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohsuho E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:111-22 (2001).
CC -!- SIMILARITY: Contains 16 Big-1 domains.
CC -!- SIMILARITY: Belongs to the intimin/invasin family.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 1315.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF005423; AAC57041.1; ;
DR EMBL; AF002559; BAB36198.1; ALT_FRAME.
DR EMBL; AF002559; BAB36199.1; ALT_FRAME.
DR InterPro; IPR003344; Big 1.
DR InterPro; IPR003535; Intimin.
DR InterPro; IPR008964; Invasin_intimin.
DR InterPro; IPR000601; PKD.
DR Pfam; PF02369; Big_1; 16.
DR PRINTS; PR01369; INTIMIN.
DR SMART; SM00634; BID_1; 16.
DR SMART; SM00089; PKD; 8.
KW Hypothetical protein; Repeat; Complete proteome.
FT DOMAIN 738 834
FT BIG-1 1.
FT DOMAIN 840 929
FT BIG-1 2.
FT DOMAIN 931 1033
FT BIG-1 3.
FT DOMAIN 1042 1132
FT BIG-1 4.
FT DOMAIN 1134 1236
FT BIG-1 5.
FT DOMAIN 1245 1335
FT BIG-1 6.
FT DOMAIN 1337 1439
FT BIG-1 7.
FT DOMAIN 1448 1539
FT BIG-1 8.
FT DOMAIN 1548 1652
FT BIG-1 9.
FT DOMAIN 1653 1750
FT BIG-1 10.
FT DOMAIN 1751 1855
FT BIG-1 11.
FT DOMAIN 1856 1957
FT BIG-1 12.
FT DOMAIN 1963 2056
FT BIG-1 13.
FT DOMAIN 2065 2156
FT BIG-1 14.
FT DOMAIN 2157 2252
FT BIG-1 15.
FT DOMAIN 2254 2355
FT BIG-1 16.
SQ SEQUENCE 2660 AA; 280062 MW; 01EB92A08F5C09D2 CRC64;
Query Match 2.2%; Score 198; DB 1; Length 2660;
Best Local Similarity 18.7%; Pred. No. 0.024;
Matches 288; Conservative 187; Mismatches 580; Indels 486; Gaps 67;
QY 228 RQINSHNPSLGNPLTWSFGYTPGK---NGILGQWITSMTPAGLKETVYNNQOG 284
Db 671 QQLNTA--VSDNVKRP-----GVTTDWKETAGDVYKATYATYKSGSLTAKLMMQWNED 723
QY 285 HHF-----POSANLPVLPVYVTLMKQVFGAGQPAIAEYSYTSNHYVGGSGNGIWNKK 336
Db 724 LHTAGFILDANPQSAKIATL-----SASNNGLANANAANTSVNVADEGSFNINDHT 776
QY 337 LD--NLYGLMTEYNGSTESRRYKQEGHDQIVRIERTYNNVHLLTSECKQONGYIQTE 394
Db 777 VTFPAVLGSGSATSFNNQNTAKTDVNGLATFD--LKSSKQEDN---TVEVTLNENGVKQT-- 828

QY 395 TAYYAIIGHNEDSOPQOLPKTKETWRSADNSYRSEITETTTDESGNPLTKVICKDKT 454
DB 829 ---LIVSFVGSSTAQVDLQSKNEV---VADGNDSSATMTATVRDAKGNLLNDV---KVT 879
QY 455 OKIISPSHWEYPPAGEVNDPCPEPYGFTFRVKKIIQTPTVDSEFFKDDPEKFIQYRSLI 514
DB 880 FNVNSAAA-----KLSQTEVNSH-----D 898
QY 515 GSQSHVTLKIERHYSATOLLNSTLFQYNTDKSELGRLLKQTECTKENGK---TYSVVH 571
DB 899 GIATATLTSKNGDYTVTASVSS-----GSAQNOQVIFIGDQSTAALTLSV-- 944
QY 572 KFTYTKQDDTLQOQSHIITHNFTHRSQVRSRYTGRFLFSDTD-----TKDIVTQMSYDK 626
DB 945 -----PSGDIIVTWTAPLHMTATLQDKNGNPLKDKKEITFSVPNDVASRFSISN 992
QY 627 LGRLLT-----RTNSGTPYANTLTYDYELNLIODNRRPFVI 664
DB 993 SGKGMDSNGTAIASLTGTLAGTHMITARLANSNVSDTQPMTF-----VADKRAVVVL 1046
QY 665 TTT--DVNGN-----QLRNEFDGAGRHVSQCLKSDGDKGFYTIHTQQVDEQGRHH 713
DB 1047 QTSKAEIIGNGVDTELLTATVKDFDNNVNLVSVFRTSPADTOL--SLNARNTNENGIAE 1105
QY 714 TS-----TYSDYLTNGRQOTD-----PDKVHLSMSKSYDNWGOIANTHMSYGVSEK 759
DB 1106 VTLKGTVLGVHTAEAILLNGNRDTKIVNIAPDASNAQVTLNIPAQQVVTNNS----- 1157
QY 760 ITVDEILTATKQLQSNVNVOTKGYTYTTPSQOPIQITLFDFA----- 804
DB 1158 ---DSVQIAT--VKDPSNHPVAGITVNFMPQDVAANFTLNNNGIAITQANGEAHVTLK 1212
QY 805 CHLOSCHTLT-----RDGMDRVKRETD-----IQOCTIY----- 834
DB 1213 GKAGTHVTATLGNNSADAQPTFVADKDSAAVVVLQTSKAEIIGNGVDTELLTATVKD 1272
QY 835 QYDYNVRVQITL---PDGTVNRKYAFSTDTLITDIRVNGISLGOOTFDGLSRLTQSQ 891
DB 1273 PFDNAVKDQVTFSTFNADTQLSQSKSN--TNDSGVAEVTFGKTVLGVHTAEA----- 1323
QY 892 DGRVWATYSAGNDQCPSTVITPDQGFIIHYQVPELDDAVLQVASNE-----ITQQ 943
DB 1324 -----TLPNGNNDKIVNIADAS---NAQVTLNIPAQQVVTNNSVQLTATVKD 1371
QY 944 FSNYPVTGALKKVAEQSLTPIYPSPGRUKMEN-----INDMKMSVLTWLRGLENGYTD 999
DB 1372 PSNHPVAGITV-----NETMPQDVAANFTLNNNGIAITQANGEAHV--TLKGGKAGTHT 1423
QY 1000 LTGTIQRKISRDTHGRVTOIKDSSIKTTLNVDLNRHIGSOVTDLATGMLTITVTFPDGLN 1059
DB 1424 VTALSNNSNDSQPTFVADKTSALVVLQISKREITGNGV--DSAT--LTATVK--DQFD 1478
QY 1060 REIGRK-----LCDSSGHTLDIQOSWLTQQLANRIVKLVNGVORTQOYSYDSNRNLNQYK 1115
DB 1479 NEVNNLVPFTSTASSGLTLPGES--NTNESGIAQAATLAGV----- 1517
QY 1116 CDGAECPDKYGHISVTONFYDIYGNITACHITFADGTEDHATFKFANPTDPCQTEVH 1175
DB 1518 -----AFGEQTVTASIA-----NNGASDNKTVHIFIGDTAAKIIEL-- 1553
QY 1176 HTHPMDPNI---RLUKYDKAGRVINIT--DNH-----GNTENFTYDTLGRLQNGQSVGY 1226
DB 1554 ---TPVDSIIAGTPQSSGSVITATVVDNNGPPVKGTVNTFNSTAATAEMTNGQAVTN 1610
QY 1227 DPLNRLYSQKTDLDLCELYRETMVNEVNGEMIRLLRGTETIIAQQRASKVLLTGTS 1286
DB 1611 EQGKATVYNNTRSSISGARPDPTVEASLENGS--STLSTSIINVNAD--ASTAHLTLLOA 1666
QY 1287 QOSVILTSKQNLQSOEAYSAYGKHKSTANDASILNGVGERADPVSQVTHLNGYRSDPT 1346
DB 1667 LFDVTSAGDITNLYIEVKDNY-----GNGVPPQOEVT 1697
QY 1347 LMRHTPDSLSPFGAGGAGINPYSYCLGDPINRSDPSGHLSHQAWTIGMGIAGLLLLIATG 1406

DB 1698 L-----SVSP--SEGTPSN-----NAYITNHGCPN-YASFTATKAGVVQVTTALENG 1743
QY 1407 GMAIAAAGIAAAIASTSTTALAFGALSVTSDITSIVSGALEDASPASKSILGWVSMGMG 1466
DB 1744 D-----SMQQTIVYVVPVNAEISLAASK----- 1767
QY 1467 AAGLAESAALKGTGLATHLGAEDGENALLKSTSESSRIKWGVTRSLDRIVRN---E 1522
DB 1768 -----DPVIANNNDUTLTTATVADTEGNAIANSE-----VTFLLPDEVDRANFTLGD 1813
QY 1523 EQQVIXDRHSRGYTDNFMKGEGQAILVHGDQDGLFVHT--EGNKHGKGPYTRHTPEQL--VD 1580
DB 1814 GKKVTD-----TEGRAKVTLKGTAG--AHTVTASMGCKS-----EQLVVN 1854
QY 1581 YLKD-----NNI-----VDLTQGGDKPV 1598
DB 1855 FIADTLTAQVNLNVTEDNFIANNVGMVTRLOATVTDGNGNPL 1895
RESULT 13
RBP2 PLAVB
ID RBP2 PLAVB STANDARD; PRT; 2867 AA.
AC Q00799; Q9N2M3;
DT 01-APR-1993 (Rel. 25, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Reticulocyte binding protein 2 precursor (PVRBP-2).
GN RBP-2 OR RBP2
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A., AND REVISIONS TO 2438-2439.
RX MEDLINE=92315338; PubMed=10838229;
RA Gallinski M.R., Xu M., Barnwell J.W.;
RT "Plasmodium vivax reticulocyte binding protein-2 (PVRBP-2) shares structural features with PVRBP-1 and the Plasmodium yoelii 235 kDa rhoptry protein family";
RT Mcl. Biochem. Parasitol. 108:257-262 (2000).
RL [2]
RN SEQUENCE OF 1189-2439 FROM N.A.
RP MEDLINE=92315338; PubMed=1617731;
RA Gallinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax merozoites";
RL Cell 69:1213-1226 (1992).
CC -!- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to human reticulocyte cells.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; AF184623; AAF76525.1; -;
DR HSP; P03069; IGCN.
KW Malaria; Receptor; Signal; Transmembrane; Repeat.
FT SIGNAL 1 21
FT CHAIN 22 2867
FT DOMAIN 22 2805
FT TRANSMEM 2806 2826
FT DOMAIN 2827 2867
FT DOMAIN 44 133
FT DOMAIN 1112 1285
FT DOMAIN 2758 2785
FT REPEAT 2758 2761
FT REPEAT 2762 2765


```

RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.,
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE OF 2-682 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92115567; PubMed=1766878;
RA Sadosky A.B., Gray J.A., Hill C.W.;
RT "The RhsD-E subfamily of Escherichia coli K-12.";
RL Nucleic Acids Res. 19:7177-7183(1991).
RN [4]
RP REVISION TO 442.
RA Hill C.W.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP REVIEW.
RX MEDLINE=95020608; PubMed=7934896;
RA Hill C.W., Sandt C.H., Vlazny D.A.;
RT "Rhs elements of Escherichia coli: a family of genetic composites
each encoding a large mosaic protein.";
RL Mol. Microbiol. 12:865-871(1994).
CC -!- FUNCTION: Rhs elements have a nonessential function. They may play
an important role in the natural ecology of the cell.
CC -!- SIMILARITY: BELONGS TO THE Rhs FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X60998; CA43309.1; -.
DR EMBL; AE000242; AAC74538.1; -.
DR EMBL; D90785; BAA15087.1; -.
DR EMBL; D90786; BAA15094.1; -.
DR EMBL; D90787; BAA15106.1; -.
DR EMBL; L19083; AAB47716.1; -.
DR PIR; C64898; C64898.
DR EcoGens; EG10010; rhsE.
DR InterPro; IPR001826; Rhs.
DR Pfam; PF03527; Rhs; 1.
DR Pfam; PF05593; Rhs repeat; 3.
DR PRINTS; PR00394; RhsPROTEIN.
DR Multigene family; Complete proteome.
SQ SEQUENCE 682 AA; 77141 MW; P704868D6E15601B CRC64;

Query Match 2.1%; Score 188.5; DB 1; Length 682;
Best Local Similarity 20.9%; Pred. No. 0.0092;
Matches 147; Conservative 83; Mismatches 212; Indels 263; Gaps 37;

Qy 884 LSLRTQSDGGRVWAYTYSAGNDQCPSTVITPDQGFTHQYQPELDDAVLQVANSNIT-- 941
Db 5 LTLTSHSTSEGRV-----SVHYGY-----DD-----KGLTGE 32

Qy 942 QQFSYNPVTGALLKAVAGQSLTIPIYPSG---RLKMNENDMKMSY-LWTLRGLNGY 997
Db 33 RQTVENPETGELLWHETGHA---YNEQGLANRVTPDSLPPVWLTLYGCVLAGMKLGG 88

Qy 998 TDLRG-TIQKTSRDTHGRVTQI--KDSSIKTLNYDDL-----NRHTGSQVTLATGHMLT 1050
Db 89 TPLEFTRDLRHRTVTSFGSMAGSNAAKLTSTYPAGQLQSOHLNSLYD----- 140

Qy 1051 TTVEFDGLNRIEGRKLCSSGHTLIDIQSWLKTQQLANRIVKLVGLQRTQEQSYDSRNR 1110

```

```

Db 141 -----RDYG-----WNDGDL-----VRIISGRQ-TREYGYSATGR 170
Qy 1111 LNQYKCDGAE-----PTDKYGHSI-----VTQN-----FTYDIYGN 1142
Db 171 LSVRTLAPDLDIRPYATDPAGNRLPDDELHPDSTLTWPDNRIAEADAHVYRDEYGR 230
Qy 1143 ITACTHTFADG---TEDHATFRFANPTDPCQLTEVHTHTPMDPNIRLKYDKAGRVINIT 1199
Db 231 LTEKTDRIIPAGVIRTDERT-----HHV-----YDSQHELVEFT 265
Qy 1200 D-NHGN---TENFTYDTLGR-----LQNGQSVVGYDPLRLVLSQK 1236
Db 266 RIQHGEPLVESRYLYDPLGRNRMAKVRERRDLTGMSLSRKRPEVTWTYGDG-DRLTTVQ 324
Qy 1237 TDTLOCELYR-----ETMLVNEVRNG-----EMIRL 1263
Db 325 TDTTRIQTVEPGSFTPLIRVETENGEREKAQRSLAETLQOEGSENGHVVFPAELVRL 384
Qy 1264 LRTGETIIAQPAS-----KVLLTGTDSQ---QSVIL 1292
Db 385 LDRLEEIRADRVSSERAWLAQCGLTVEQLARQVEPYTPARKVHFYCHDRGLPLALI 444
Qy 1293 TSDKQNLQSEAYSAVKEHKSTANDASI---LGYNGERADPVSGVTHLGNVRSYDPTLMR 1349
Db 445 SEDGNTANRGEYDEWGNQNLNENPHLHPQYRLPQQQHDSESLY--NRHRHYDPLQCR 502
Qy 1350 FHTPDSLSFGF-AGGINPYSYCLGDPINRSPGSHLSQAWTGMGTAGIALLTIATGGM 1408
Db 503 YITPD---PIGLRGGMNYYPL-NPIQVIDPMG-----LDAIENNTSGGL 544
Qy 1409 ATAAAGGIAAATASSTP--ALAFGALSVTSDITSIVSGALEDAS 1451
Db 545 -IYAVSGVPLIANSITNSAYQFG-----YMDAIVGGAENGAA 583

RESULT 15
ID YDBA ECOLI STANDARD; PRT: 2003 AA
AC P33666; P76087; P76088; P76856; P76857; P76859;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ydba.
GN YDBA OR B1401/B1405.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [3]

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2004, 05:45:33 ; Search time 68 Seconds
(without alignments)
7762.676 Million cell updates/sec

Title: US-09-889-874A-23

Perfect score: 8879
Sequence: 1 VIKELKLFRRITMSDNEF.....PKIILGTEKTKVTKTRP 1673

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp Vertebrate:*

14: sp Unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8815	99.3	1660	2 Q9EVR7	Q9evr7 xenorhabdus
2	1614	18.2	1562	16 Q883V9	Q883v9 pseudomonas
3	1553	17.5	773	2 Q8S6T1	Q8s6t1 coxiella bu
4	1530.5	17.2	774	2 Q52880	Q52880 coxiella bu
5	1416	15.9	1632	16 Q88LP7	Q88lp7 pseudomonas
6	1214	13.7	709	2 Q45948	Q45948 coxiella bu
7	1037	11.7	1290	16 Q88LP2	Q88lp2 pseudomonas
8	932.5	10.5	982	16 Q883W6	Q883w6 pseudomonas
9	920.5	10.4	1669	16 Q87VG6	Q87vg6 pseudomonas
10	919.5	10.4	505	2 Q45905	Q45905 coxiella bu
11	919.5	10.4	526	2 Q9X626	Q9x626 coxiella bu
12	916.5	10.3	526	2 Q52883	Q52883 coxiella bu
13	912.5	10.3	528	2 Q45949	Q45949 coxiella bu
14	884	10.0	589	2 Q45902	Q45902 coxiella bu
15	648	7.3	762	2 Q93IJ8	Q93ij8 vibrio sp.
16	478.5	5.4	310	2 Q45904	Q45904 coxiella bu

17	450.5	5.1	2217	17	Q8TP72	Q8tp72 methanosarc
18	429.5	4.8	336	2	Q45946	Q45946 coxiella bu
19	420.5	4.7	2364	16	Q82R58	Q82r58 streptomyc
20	395	4.4	820	16	Q82QR1	Q82qr1 streptomyc
21	393.5	4.4	1976	16	Q8D4R5	Q8d4r5 vibrio vuln
22	392	4.4	1250	16	Q82RX1	Q82rx1 streptomyc
23	391.5	4.4	2370	16	Q82RE3	Q82re3 streptomyc
24	388.5	4.4	1826	16	Q987Z7	Q987z7 rhizobium l
25	386	4.3	2167	16	Q92EK5	Q92ek5 listeria in
26	386	4.3	2224	16	Q81U00	Q81u00 bacillus an
27	382.5	4.3	1117	16	Q8CK70	Q8ck70 streptomyc
28	381	4.3	2183	16	Q86585	Q86585 streptomyc
29	376.5	4.2	2082	16	Q9ACP4	Q9acp4 streptomyc
30	365	4.1	1400	16	Q8X2F8	Q8x2f8 escherichia
31	362	4.1	1411	2	Q46748	Q46748 escherichia
32	361	4.1	1510	16	Q8PHK6	Q8phk6 xanthomonas
33	359	4.0	380	16	Q87UI5	Q87ui5 pseudomonas
34	354	4.0	1710	16	Q8XTG0	Q8xtg0 ralstonia s
35	353.5	4.0	332	16	Q88AD8	Q88ad8 pseudomonas
36	353.5	4.0	843	16	Q82950	Q82950 salmonella
37	352.5	4.0	1410	2	Q52673	Q52673 escherichia
38	348.5	3.9	1394	2	Q52668	Q52668 escherichia
39	348	3.9	1381	16	Q87PI5	Q87pi5 vibrio para
40	341	3.8	1394	16	Q8X385	Q8x385 escherichia
41	340	3.8	1316	16	Q8A692	Q8a692 bacteroides
42	338	3.8	1426	2	Q93V17	Q93v17 escherichia
43	338	3.8	2306	16	Q8F107	Q8f107 leptospira
44	337.5	3.8	2321	16	Q8F5B9	Q8f5b9 leptospira
45	336.5	3.8	1404	16	Q8XED9	Q8xed9 escherichia

ALIGNMENTS

RESULT 1

Q9EVR7 PRELIMINARY; PRT; 1660 AA.
ID Q9EVR7;
AC Q9EVR7;
DC 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Nematocidal protein 2.
EN KNP2.
OS Xenorhabdus bovienii.
GN Xenorhabdus bovienii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Xenorhabdus.
OX NCBI_TaxID=40576;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=I73;
RA Morgan J.A.W., Ousley M., Ellis D., Jarrett P.;
RT "Novel toxins from Xenorhabdus strains with activity against nematodes."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ296651; CAC19493.1; --
DR InterPro; IPR006530; YD.
DR Pfam; PF05593; RHS_repeat; 5.
DR TIGRFAMs; TIGR01543; YD_repeat_2x; 12.
SQ SEQUENCE 1660 AA; 185738 MW; AFABEA20AD70B164 CRC64;

Query Match 99.3%; Score 8815; DB 2; Length 1660;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1660; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 MSDNNEFTQANNFTSAVSGGVDPRGTGLYNTQITLGHIVGNGNLGPTPLTILSYSPLNKT 73

Db 1 MSDNNEFTQANNFTSAVSGGVDPRGTGLYNTQITLGHIVGNGNLGPTPLTILSYSPLNKT 60

QY 74 DIGFGIGNFGLSVYDRKNSLLSLSTGENYKVIETDKTKLQOKKLDNLRTEKDKENY 133

Db 61 DIGFGIGNFGLSVYDRKNSLLSLSTGENYKVIETDKTKLQOKKLDNLRTEKDKENY 120

QY 134 RIHKSGDIEVLGTGFNNNAFDLKVPKLLNFAGHAIYIDWNFEATQPRNLRIYDDLDGDH 193

Db 121 RIIHSGDIEVLGTGNNNAFDLVKPKLLAPAGHAIYIDWNFEATQRLNRIYDDLDGHD 180
Qy 194 IPLNLEYQGLIKITLTLFPQKEGYRTELRFNLNQLNSIHNFSLGNENPLTWSFGYTP 253
Db 181 IPLNLEYQGLIKITLTLFPQKEGYRTELRFNLNQLNSIHNFSLGNENPLTWSFGYTP 240
Qy 254 GKNGILGOWITSMAPGLKETVNYSNNOCHHFPQSANLPVLYPYTLKMKVQAGOPAI 313
Db 241 GKNGILGOWITSMAPGLKETVNYSNNOCHHFPQSANLPVLYPYTLKMKVQAGOPAI 300
Qy 314 QAEYSYTSNHYVGGSGNGIWNKLDNLYGLMTEYNYGSTESRRYKQKEGHDQIVRIERTY 373
Db 301 QAEYSYTSNHYVGGSGNGIWNKLDNLYGLMTEYNYGSTESRRYKQKEGHDQIVRIERTY 360
Qy 374 NYHLLTSECKQNGYIOTTETAYAIIGHNFDQSOFQPKTKTETWRSADNSYRSEI 433
Db 361 NYHLLTSECKQNGYIOTTETAYAIIGHNFDQSOFQPKTKTETWRSADNSYRSEI 420
Qy 434 TETTFDSGNPLTKVKKDKTKTKIISPSHWEYYPAGEVDNCPBPYPGFTFRVKKIIOT 493
Db 421 TETTFDSGNPLTKVKKDKTKTKIISPSHWEYYPAGEVDNCPBPYPGFTFRVKKIIOT 480
Qy 494 PYDSEFKDDEKFTQYRYSIGQSHVTLKIEBRHYSATOLLNSTLFOYNTKSELGRLL 553
Db 481 PYDSEFKDDEKFTQYRYSIGQSHVTLKIEBRHYSATOLLNSTLFOYNTKSELGRLL 540
Qy 554 KQTECTKGNGKTVSVVHKFTYKQDDTLQOSSHITTHDNFTTHRSQVRSRYTGRLPST 613
Db 541 KQTECTKGNGKTVSVVHKFTYKQDDTLQOSSHITTHDNFTTHRSQVRSRYTGRLPST 600
Qy 614 DTKDIVTQMSYDKLGRLLTTLNSGTPYANTLYDYVELANNLQDDNRPFPVITTDVNGNQ 673
Db 601 DTKDIVTQMSYDKLGRLLTTLNSGTPYANTLYDYVELANNLQDDNRPFPVITTDVNGNQ 660
Qy 674 LRNEFDGAGRHVSQCLKSDGCKEFTYHTQOYDEQGRHHTSYSDYLTNGRCQDTPDKV 733
Db 661 LRNEFDGAGRHVSQCLKSDGCKEFTYHTQOYDEQGRHHTSYSDYLTNGRCQDTPDKV 720
Qy 734 HLSMSKSYDNWGIANTHWSYGVSEKITVDPILTATKQLQSNNSNVQTKVTTVTPSQ 793
Db 721 HLSMSKSYDNWGIANTHWSYGVSEKITVDPILTATKQLQSNNSNVQTKVTTVTPSQ 780
Qy 794 OPTQITLFDAGHLSQCHTLTRGWRVRKETDAIGCCTIYQVDNVRVIOITLPDGTIV 853
Db 781 OPTQITLFDAGHLSQCHTLTRGWRVRKETDAIGCCTIYQVDNVRVIOITLPDGTIV 840
Qy 854 NRKYAPSTDTLITDIRVNGISLQQTFDGLSRLTOSQDGRVWATYVAGNDQCPSTVI 913
Db 841 NRKYAPSTDTLITDIRVNGISLQQTFDGLSRLTOSQDGRVWATYVAGNDQCPSTVI 900
Qy 914 TPDQCFHIYQBELDAVLQVANSNEITQOFSNPVTGALLKAVAGQSLTIPIYPSGR 973
Db 901 TPDQCFHIYQBELDAVLQVANSNEITQOFSNPVTGALLKAVAGQSLTIPIYPSGR 960
Qy 974 KMENINDMKMSYLWLRLGNGYDTLTGTIQKISRDTHGRVTOIKDSSIKTLLNVDL 1033
Db 961 KMENINDMKMSYLWLRLGNGYDTLTGTIQKISRDTHGRVTOIKDSSIKTLLNVDL 1020
Qy 1034 RHIGSQVTDIATGHMLTTTVEPGLNREIGRKLCDSSGHTLDIQOSWLKTOQLANRIVKL 1093
Db 1021 RHIGSQVTDIATGHMLTTTVEPGLNREIGRKLCDSSGHTLDIQOSWLKTOQLANRIVKL 1080
Qy 1094 NGVLQRTQESYDSRNLNQYKCDGACPTDKYGHISIVTQNFYDLYGNITACHTTFADG 1153
Db 1081 NGVLQRTQESYDSRNLNQYKCDGACPTDKYGHISIVTQNFYDLYGNITACHTTFADG 1140
Qy 1154 TEDHATFKFANPTDPCQLTEVHHTHPMDPNILKYDKAGRVINITDNHGNTEFTYDTL 1213
Db 1141 TEDHATFKFANPTDPCQLTEVHHTHPMDPNILKYDKAGRVINITDNHGNTEFTYDTL 1200
Qy 1214 GRLONGGSSVGYDPLNRLVSQTDTLDCELYRETMLNVRNGEMIRLLTGTETIAQ 1273

Db 1201 GRLONGGSSVGYDPLNRLVSQTDTLDCELYRETMLNVRNGEMIRLLTGTETIAQ 1260
Qy 1274 QRASKVLLTGTDSQOSVILTSDKQNLQOEAYSAYGKHKSTANDASILGYNGERADPVSGV 1333
Db 1261 QRASKVLLTGTDSQOSVILTSDKQNLQOEAYSAYGKHKSTANDASILGYNGERADPVSGV 1320
Qy 1334 THLNGVYSYDPTLMRPHTPDLSLPGAGGINPVSICLGDPINRSDPSGHLSSQAWTIG 1393
Db 1321 THLNGVYSYDPTLMRPHTPDLSLPGAGGINPVSICLGDPINRSDPSGHLSSQAWTIG 1380
Qy 1394 MGIAGLLTIATGGMAIAAAGGIAAAIASTSTTALAFGALSVTSDITSIVSGALEDASP 1453
Db 1381 MGIAGLLTIATGGMAIAAAGGIAAAIASTSTTALAFGALSVTSDITSIVSGALEDASP 1440
Qy 1454 ASSILGWYSMGAGLAESAIGKGTKLATHLGAFAEDGENALLKSTSSSRIKWGYTRS 1513
Db 1441 ASSILGWYSMGAGLAESAIGKGTKLATHLGAFAEDGENALLKSTSSSRIKWGYTRS 1500
Qy 1514 LDREIVRNEEGQVTKDHSRGYTDNFMGKGEQAILVHGDKGFLYHTEGNKHNGKGPYTRH 1573
Db 1501 LDREIVRNEEGQVTKDHSRGYTDNFMGKGEQAILVHGDKGFLYHTEGNKHNGKGPYTRH 1560
Qy 1574 TPEQLVDYLDKNNIVDITQGGDKPVHLLSCYKSGSSGAADKMAKYNRPVAYSNKPTISQ 1633
Db 1561 TPEQLVDYLDKNNIVDITQGGDKPVHLLSCYKSGSSGAADKMAKYNRPVAYSNKPTISQ 1620
Qy 1634 GLARIERKDFLTKSYHSYDPRKILGRTEKTVKPKTFRP 1673
Db 1621 GLARIERKDFLTKSYHSYDPRKILGRTEKTVKPKTFRP 1660

RESULT 2
Q883V9
ID Q883V9 PRELIMINARY; PRT; 1562 AA.
AC Q883V9
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE YD repeat protein.
GN PSPT02239.
OS Pseudomonas syringae (pv. tomatop).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Neilson W., Davidson T.,
RA White O., Fraser C., Collier A.;
RT "Complete sequence of Pseudomonas syringae."
RT Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016863; AA055755.1; -
DR TIGR; PSPT02239; -
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR GO; GO:0006310; F:DNA recombination; IEA.
DR GO; GO:0006281; F:DNA repair; IEA.
DR GO; GO:0006260; F:DNA replication; IEA.
DR InterPro; IPR000977; DNA_ligase.
DR InterPro; IPR006530; YD.
DR Pfam; PF05593; RBS repeat; 6.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW Complete proteome.
SQ SEQUENCE 1562 AA; 175713 MW; 8DC10DA1BFE37BF1 CRC64;
Query Match 18.2%; Score 1614; DB 16; Length 1562;
Best Local Similarity 30.3%; Pred. No. 1.4e-78;
Matches 490; Conservative 239; Mismatches 700; Indels 190; Gaps 51;
Qy 14 MSDNNEFFTOANNFTSAVSGVDPRIGLYNIQITLGHIVGNLGLPTLPLTSLSPLNKT 73

Db 1 MTSTSVHNAFMYSYVLSQGVDPRTGQYTVSNLPEVKSNGLRGPVVLVLSYNPLNVQ 60
Qy 74 DIFGFGFNPLSVYDRKNSLLSLSTGENYKVI-ETDKTVKLOQKKLDNLRFEKDLKENC 132
Db 61 DSGFGLGWNLLQSDYDPGTRIVSLGSGCTFKVEGSLGDQLMPEKKLDSHFYKQ-DDTR 119
Qy 133 YRIIHKSGDIEVLTFGNNAFDLKVPKLLNPAGHAIYIDWNFEATQPSLNIYDDLDGH 192
Db 120 YRVHKSQGVVEEVLVLSGNLRPALVRIYSPGEGHITLHYASFGAYQMLSEWDD-DQO 178
Qy 193 DIPLLNLEYOGLIKTILTFPGQKEGYRTELRFNLQNSIHNFSNGENPLTWSGYTP 252
Db 179 VILITIRD-----STSVRLLLYGAPKADAEPVILSGSNENVARIELFTANKASWRETYGI 234
Qy 253 IGKNGILGOWITSWTAPGLKETVYNNQGHFFQSQANLPLPVYVTLMKQVPGAGQPA 312
Db 235 IRGHSC-----IASVDPVVGHEVDVYQDS--GHQFLLSAGREPLRVTHLITPGFLOPE 288
Qy 313 IQAEYSY-----TSHNVVGGSGNIW--NNKLDNLGLM--TEYNGSTESRYKDKEGHD 364
Db 289 VDVRAYKDCAGRERNFLGAGLDIAWEDNGLDNLRYLGAAPLYLSTETLRVNDVD--- 345
Qy 365 QIVRIERTYNNYHLLTSCQOQNGYIQTETAYYAIIGHNFDSPQSQFOLPKTKTETWR- 423
Db 346 -VRSIERVFNOFHLLAETTRONLSILEVDTRYIEBEGKFFDLQPNYCOLPKREVRTWL 404
Qy 424 SADNSY-RSITETTFDESNGPLTKVIKDKTKIISPSHWEYPPAGEVDNCPPEPYG 482
Db 405 SPGSGVPTIIVSDYSDYNLLAQAQANGVETS-----EWYSVSGE-DCGPPDPG 456
Qy 483 FTRPVKKIITPVDSSEPKDPEKPIQVRY-----SLIGSQSHVTLKIE-----BRHYSATQ 533
Db 457 FVRTLAKSVVPAQSDYGHALVLTTRYRYKALPALAGSGONLWAAESETLLQQTIDGK 516
Qy 534 LLNSTLFOYNTDKS---ELGRLLKQTECKGNGKT---YSVVHKTFTYTKODTLQOQH 586
Db 517 ELQOTTYYIEDNPDAFQGRIRHQSVTLEGSLTTDYRYDLQDPDF---DQTVQQTIV 573
Qy 587 SIITHDNFTHRSQVR---SRYTGRLEFSDTDTKDVIQMSYDKLGRLLTTLN-SGTPYA 642
Db 574 QIVT--GFDMTQKVRLEHSLFTGCEPLNDRDNDVEIRYDYLRLVRVSETVSPKEEYK 631
Qy 643 NLTLYDELNNLQDNDPPFVIITTDYNGQLRNEPDGAGRHSYQCLKSDGDKGYTTH 702
Db 632 ATRHYEYQLCAVKTDQAEQRLF---DVKNVQTSRFDGLGRVIEARADADNPDVHRRLD 688
Qy 703 TOQ-----YDEQGRHHTSTYSYLTLNGRQQTDPKVLHLSMSKSYDNWQGIANTHWSYGS 757
Db 689 LRQYEAAYDAWQKVBETSVDWL-----DQKRALTYVEYDDWDQQLSVTGPDGVT 741
Qy 758 EKITVDITLITAT-----KQLQSNNSNVQCKEVITY-TPSQQPIQITLFEAGHLQSCH 811
Db 742 TIEQTDVPGTQASNGPQRRWTESNDGLQTSVSETWNLNLFDEPTESVRLDRDLWSEPV 801
Qy 812 TLTR--DGMVRVAKETDAI--CQCTIYOYDNNRVYQITLPGTIVNRYKAPFSDTL 865
Db 802 SLRSYQVGLGRVKEVSGCLPIRESITYGYDVEDRVANTLPGAVVRRYAPHSGEDL 861
Qy 866 ITDIRV--NGIS--LGOQTFDGLSRLTQSQDQGGRWAYTYAGNDQCPSTVITPDGQFIH 921
Db 862 PAWIGVDHNGKSVLGEQKFGDLRIVTSITGGRERELSVTSDLMQ-PKTVKLPSGRQID 920
Qy 922 YQYQPELDDAVLO-----VASNEITQOFSNPNVYTGALLKAVAGSGLTPIYVPSGRKLM 975
Db 921 YDYLPELGEDEPKTKQSDTVARLKLTADYITDPQARLIGSSEGEELREYVSTGSLKS 980
Qy 976 ENINDM---KMSYMLWTLRGLNGYVTLTGTIQKISRDTGHRVTQIKDDSIKTLNYYDDL 1032
Db 981 EORTSQGLIENTHMYRSLGLPLSVLDVLGQQLSVYDDFGLAQTSLGGEVVSFTYDTF 1040
Qy 1033 NRHIGSQVTLATGHMLTTTTFEPLGLNREIGRKLCDSGHTLDTQCSWLKTOQLANIVK 1092

Db 1041 CRTASITLDSSSNGQVVVISLEYDAQREAQRTF-TINGANQQMVQYVDDVDQVQKXTLS 1099
Qy 1093 LNVQLORTEQYSYDSRNLNOYKCDGAECPDPTKYGHSIVITQNTFYDIYGNITACHTTFAD 1152
Db 1100 EGAVIIEEHYCYDLGRLTYQDCSGKQRPVDPYGMTISRQVFSFQDGLNLLTTLVTF-D 1158
Qy 1153 QTEPHATFKFANPTDPQOLTVEVHHT-----PMPDNIRLKYDKAGRVIN 1197
Db 1159 GGRNRARY-FYEGIDPAQLTRVINTQLAWNNARLIPVKNKNDSTYPPPEIRLTYDPGDL-- 1215
Qy 1198 ITDNHGNTEYTYTLGRL---QNGQGSV-YGYDPLNRLVSQKTDTLDCELYVRETMLVN 1253
Db 1216 ITDEADRL--JUSDPLGRLLEVSNPAGVRYRYPQDELACETGE---QRFYRDGVLAS 1269
Qy 1254 EVRNGEMIRLLRTGETIIAQOASKVLITGTDQSOVILTSKONLSOEAYSAYGKHKST 1313
Db 1270 QLGASQNSTYWKRGDGYLLAEQGGSDALLFSTISNSVLSVSEVHPDGVSNRSYTVYGHSSGD 1329
Qy 1314 ANDASILGYNGERADPVSGVTHLNGYRSYDPTLMRHTPDSLSPFGAGGNNPYSCIGD 1373
Db 1330 DPFAGRLYNGELHETDTGWLQGLNGYKAYNPVLMRFHSPDSWSPFFGEGGJNAYAGEGD 1389
Qy 1374 PINRDPDPSGH-----LSWQAWTG-----IGMGIAGLLLTATG 1407
Db 1390 SVNGVDPTGHWFKLPRIFRTRRLKQAKADKKEFWRLIEQDIKNEGLEGRYAQAYRDL 1449
Qy 1408 MAIAAAG---IAAAIATSTTALAFGALSVTSDITSVSGALEDAS--PKASSILGWVS 1462
Db 1450 QAKSANNKAYISSLVKHRREKALE--AETLKANTGDMSSSREESTITPKAESGSGSIS 1507
Qy 1463 MCMGAAGLAESAIGKGTGLATHLGAFABEDGENALIKSTSESRKIKGWVTRSLDRIVRN 1521
Db 1508 -----NLGFGSGDRDVGRLNVSSA---KSEVLRRQD--FVRN 1540

RESULT 3
Q9S6J1
ID Q9S6J1 PRELIMINARY; PRT; 773 AA.
AC Q9S6J1
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
DE Hypothetical protein.
OS Coccidia burnetii.
OC Plasmid QpV.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coccidiaceae; Coccidia.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1140;
RA Radonski K.U., Willems H., Lautenschlaeger S., Jaeger C., Baljer G.;
RT "sequence of QpV plasmid."
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF131076; AAD33495.1; -
DR GO; GO:046821; C:extrachromosomal DNA; IEA.
DR InterPro; IPR006530; YD.
DR Pfam; PF05593; RHS repeat; 6.
DR TIGRfams; TIGR01643; YD repeat_2x; 8.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 773 AA; 83759 MW; 4C5FCE0481CCAC44 CRC64;

Query Match 17.5%; Score 1553; DB 2; Length 773;
Best Local Similarity 44.6%; Pred. No. 9.6e-76;
Matches 342; Conservative 119; Mismatches 275; Indels 30; Gaps 17;

Qy 741 YDNWGIANTHWSYGVSEKIIVDTPLTLATKQLQSNNSNVQGVKVVTPSQQPIQITL 800
Db 4 YDSWGQNLTVFSDGQERSVYDPIRRAT--LQESGSKLQGGQTEYNLAGLIKVTQ 61
Qy 801 FBEAGHLQSCHTLTRDGNDRVRKETDAIGCQCTIYOYDNNRVYQITLPGTIVNRYKAPF 860
Db 62 YDSQGTQESAHEVYDGLQRLKETDELQGVTLVEYDFGRVTVTTLPTENTIIQKSYAPH 121

OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
 CC Cocciliaceae; Cocciliella.
 CX NCBI_TaxID=777;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=Nine Mile phase I;
 RA Thiele D., Willems H., Haas M., Krauss H.;
 RT "Analysis of the nucleotide sequence of the cryptic plasmid
 RL Qphi from Cocciliella burnetii";
 RL Eur. J. Epidemiol. 10:413-420(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Nine Mile phase I;
 RA Thiele D.;
 RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X75356; CAA53129.1; -;
 DR PIR: S38241; S38241.
 DR GO: GO:0046821; C:extrachromosomal DNA; IEA.
 DR InterPro: IPR006530; YD.
 DR Pfam: PF05593; RHS_repeat; 8.
 DR TIGRFAMs: TIGR01643; YD_repeat_2x; 8.
 KW Plasmid.
 SQ SEQUENCE 709 AA; 79417 MW; A377B9A2A72CCF86 CRC64;

Query Match 13.7%; Score 1214; DB 2; Length 709;
 Best Local Similarity 40.8%; Pred. No. 2.1e-57;
 Matches 287; Conservative 101; Mismatches 280; Indels 36; Gaps 16;

QY 554 KQTECTGKNGKTSVVKHFKYTKQDDTLOQSHSITTHDNFTTHRSOVRSGRYGRFSDT 613
 DB 3 RRAEVLTSKGGKYNQNTTFAISSQAEHLQKIDFTGGDGIKTSISREQSYSGRLLSST 62
 QY 614 DTKDIVTQMSYDKLGRLLTTLN-SGTPYANTLYTYELNMLNLODDNRPPIVITTDVNGN 672
 DB 63 DELGNVTQYDEYELGRLLTQTVNASSTYASTRYSYSLTDARGKVTAIKTTVTPDKGN 122
 QY 673 QLRNEFDGAGRHVSQCLKSDG-----DGKPYITHQYDEQGRHHTSTYSYDLTNGROQ 727
 DB 123 QLRYYDGLGNLQERLKDAAVSQKTYWYITHQQQYDALGRESKITIQDVL-----R 177
 QY 728 TDPKRV-----LSMK--SYDNMGQIANHWSGVSEKIITVDPIITATKQLQSNVNVQ 781
 DB 178 LDSEVGHAGSVLSTKVMHYDSWCQNLTVFSDGYQERSYVDPIITRAT--LQPSGSQK 235
 QY 782 TGKEVNTYPSQOPIQTLTDEAGHLOSCHTLTRDGDWRVRKETDAIGQCTIYQYDYNR 841
 DB 236 LGQQLTEYNLAGLPKIVQYDSQSGTEGSAHYFDYGLGQURKETDELQITLLEYDHFGR 295
 QY 842 VIQITLPGDTIVNRKAPFSTDLITDIRVNGISLQOQTFDGLSLRSLQDQGRVWAYTY 901
 DB 296 VTQTLPTENTIIQSYAPHSASTLITGIVSNVTSMGNTQFDSLERLTTSQGRTSAFSY 355
 QY 902 SAGNDQCPSTVITPDGFIHYQPELDDAVLQVANSNEITQPSYNPVTCALLKAV-ABG 960
 DB 356 ENAS-SVPAAVTAPGTSTVSEYELKEIGNAVKISAPILQITWDYDALGTAMTSATQAG 414
 QY 961 QSLTPIYPSGRKME-NIND---MKMSYLWTLRGLENGTYDLTGTIQKISRDTHGRVT 1016
 DB 415 MIRQMTYPSGLLKNETSPDGAQAQSTAVTYSLAGAPQSYTDVFGVTQRYDYDEHGRRI 474
 QY 1017 QIKDSSITKTLNVDLNRHIGSQVTLATGCHMLTTTVEPCLNREICRKLCDSSGHTLDI 1076
 DB 475 GIEDNDIKSLVDYDAFGRFTKQQAATDKKGAIVSTLTLYDNLNREIKREISAGQSVLVI 534
 QY 1077 QOSMLKTLQALNRIKVLNGLVQTEQVSYDSRNRLNQYKCDGACPTDKYGHISVTQNF 1136
 DB 535 EQIVQRNHLKERITQSGRTTLAKEMFAYDSRNRLIEYTCNGEARQDPVGRKAIHQTS 594
 QY 1137 YDIYGNITACHTTFAQTEDEHATFKFANPTDPCQLTEVHTHTPDMDPNILKYDKAGRVI 1196
 DB 595 YDALGNMTKTQTFSGG-RNTATYIY-SAIDPTQLLKVNNDHSDYKPEITLEYDKAGRM 652
 QY 1197 NITDNHGNTEFTYDTLGRILQ--NG---QGSVYGYDPLNRLVSQ 1235

DB 653 R--DEAGRT--LRYDALGRLLQVNGAGAKGGQYAYDVLNLTLSQ 692

RESULT 7
 Q88LP2 PRELIMINARY; PRT; 1290 AA.
 AC Q88LP2;
 DT 01-JUN-2003 (TREMELrel. 24, Created)
 DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN PP1887.
 OS Pseudomonas putida (strain KT2440).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=160488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22423060; PubMed=12534463;
 RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
 RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
 RA Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,
 RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,
 RA Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzar A.,
 RA Utterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,
 RA Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,
 RA Kiewitz C., Eisen J., Timmis K.N., Duesterhoeft A., Tuenmler B.,
 RA Fraser C.M.;
 RT "Complete genome sequence and comparative analysis of the
 RT metabolically versatile Pseudomonas putida KT2440";
 RL Environ. Microbiol. 4:799-808(2002).
 DR EMBL: AE016781; AAN67506.1; -;
 DR TIGR; PP1887; -;
 DR InterPro: IPR006530; YD.
 DR Pfam: PF05593; RHS_repeat; 6.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 1290 AA; 146578 MW; 7081A3F10B287F92 CRC64;

Query Match 11.7%; Score 1037; DB 16; Length 1290;
 Best Local Similarity 27.7%; Pred. No. 2.2e-47;
 Matches 375; Conservative 196; Mismatches 578; Indels 204; Gaps 52;

QY 14 MSDNNEFTQANNFTSAVSGGVDPRTGLYNIQITLGHIVGNLGLPTLPLTSLSPKNT 73
 DB 2 MSTSTSVHSAFNFSSYIESGVDPRTGYTVSIRLPELQNDLQPGFELALFSPLN 61
 QY 74 DIGFIGNFGLSYVDKNSLLSSTGENYKVI-ETDKTVKLOKLDNLRFKDLKEN- 131
 DB 62 DSGFGKWN--LQTLQVRKHIVTLSSGETYKITKSSVTGLENKQKLQF--DLYEDP 117
 QY 132 -----CYRIHKSQDIEVLTGFNNNAFDLVKPKLLNPAGHAIYIDW-NFEATQRLN 183
 DB 118 PPGGAARFRVHRHSGLVLEWVGSGEVALVELHSLPLGHLHLISYLPFGGHRLS 177
 QY 184 RIYDDLDGHDIPLNLEYQGLIKTILTLFPQKQEGYTEIRF---LNRQLNSIHNFSLGN 240
 DB 178 EVQDESD---VLLRLQSRNSRVLLCYPSGGDGGFLARYAMTLESNRYSEILPT 233
 QY 241 ENPLTWSFGYTPIGNKGLGQ-WITSMTPAGLKETVNSNNNGHFFPOSANLPV-LPY 298
 DB 234 ANQARWRTY-----EDVLGYLCVRECTPYGGYRVFY--QDAGHKFPSSAARDKNLPR 286
 QY 299 VTLMKVPGACQAPQIAQAEYSY-SHNYVGGSGNINWKK-LDNLGLMTYNYGSTESRR 356
 DB 287 VTRHEIDPRFQAAKVVRYEYPGTHNFIQGGSSISWSDGLDNLKYVPEDYTKSIQI-- 344
 QY 357 YKDEGHDOIVR-TERTYNNVHLITSCKQNGVIQITETAYYAIIGHNFDSPSQFQLP 415
 DB 345 ---QEVRRQSVRTITRINRPHLLTEQATQCKLQAFTRYADNAG-NFESSEPYQLP 400
 QY 416 KTKTETWRSADN--SYRSEITETTFDESGNPLTKVTKDKTKQIKISPTHEWYPPAGEV 473

Db 401 HDSTQWLSISPSRQREKRVTRYDSHGVLTRLLPNQVLETV-----WYSABERG 453
QY 474 DNPPEPYGTRFVK-KIIQ-----TPYDBEFKDDPEKFIQYV 511
Db 454 DE-----HGFVRLKTRTVQPAATGAGAAATLQAFQYRALTPGSG-----YLKQW 500
QY 512 SLIGSQSHVTLKTEERHYSATQILN--SLIFQ-YNTDKSELGLLKTQECTKG-ENGKTY 567
Db 501 RLLESET-----LSEGSAPGTLEKISKLYQESAEBSYSGVRQOTVSYGVGCGSPF 555
QY 568 SVVHKFTYTKQDD--TLQQSHSTTHDNFTIHSQVRSRYTGLRFDSDTDTKQIVTQMSYD 625
Db 556 DTLTHYGSLPDDERALTQVEXLVGVGSGESKTLTLRHALDTGEPLLNLDNGVEIRYYD 615
QY 626 KLGRLTRTLNGSTPVANTLTYDELNNLQDNRPFVITTDVNGQLNRFDFGAGRV 585
Db 616 ALARVTRVAVPKPFEAERNYKFLCAYDNEQAQQWYDKQVETHL---LDGSRPI 672
QY 686 SOCLKSDG--DGKFTYHTQOYDEGRHHTSYDYLTNGRQOTDPDKVHLSMSKSYD 742
Db 673 FEERAKDSATYAGALRPYRKYDELDQVLETEIDWLGDG-----LLELTSHISYD 725
QY 743 NWGO---IANTHSYGVSEKITVDPIILT-----ATKQOSNNNVQTKVTTTPSQ 793
Db 726 DWGRYAVLPDGTQVEE---IDQVASTDGPVHRAWREVE---HSRVSGITETWNLFE 779
QY 794 OPTQITLDFEAGHLQ-SCHTLTRDGDWRVKEHTDAIG---OCTIYOYDNNVRIQITLDP 849
Db 780 KPVRIERFALDGTSTISLQWNVYDGLRLSREAGGAGORVVEYRIFADPRVERLAD 839
QY 850 G-TIVNKYAPFSTDLITDIRV-----GISLQOQTFDGLSKLTQSDQGRVWAYTSA 903
Db 840 RVNTVRYTGAHSRNDLPVSIKVGNTESAVALLEGEQVDFGLERRITVAITGREGTPEYDP 899
QY 904 GNDQCPSTVITPDGOFHYQYQPEL--DDAVLOVASNEITQOFSYNPVTKALLKAVAGQS 962
Db 900 GERQ-PHWKAPDGTRIEYOYRPAIGEEPVLRLVSGK-EAKYBYDLKNAKLTHCEBPGDD 957
QY 963 LTPYI-----YPSGRKMKEN---INDMKMSYLVTLRLGLENGYDLDLTIQIKSRDTH 1012
Db 958 ENSGYTLDRSHFLSNGEVEKESRTVDGEAFSMYDYSFRSLRAYVDVLQOTQLYDFDDV 1017
QY 1013 GRVTOIKDSS-----IKTLYDDLNRRHIGSQVDTLACHMLTTT 1052
Db 1018 GRLEKTTLHAPEKPRTRYLRQAPARQLLLESTFGYDQGRVASITTTDASTGHATL 1077
QY 1053 VEPGLNREIGRKLCDSSGHTLTIQOSWLKTQOLANRIVK-----LNGVLQRTQOYS 1104
Db 1078 LEYDEFREILRTF-DFGDTVQTLAQDYDEDFCLKSRILKERPKGSDSQAILLRHETQ 1136
QY 1105 YDSRNLNQKCDGAECPDKYGHSVITQNFYDIYGNITACHTTFADGTEDHATPKFAN 1164
Db 1137 YDRGRQLIVTCDPGEAPVDPGQTIARQIFGFDGLNLSLVITYRDPGSGWQRTLYEFKN 1196
QY 1165 PTDPQ-----LTVVHTHPDMPD-----NRLKYDKAGRVINITHGN 1204
Db 1197 -SDPAQMSRIIPDALLTVDVHDLEELFKIVKYPQLAIDLHYDNGNL---ISDEQGR 1253
QY 1205 TENFTYDTLGRL-----QNGOGSVYGYDPLNRL 1232
Db 1254 V--LTYDGLNRLLAVEIPDGERCRNYNDPENIL 1284
RESULT 8
Q883W6 ID Q883W6 PRELIMINARY; PRT; 982 AA.
AC Q883W6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE YD repeat protein.
GN PSPTO2231.
OS pseudomonas syringae (pv. tomato).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
RN NCBI_TaxID=323;
RP [1]
RC SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Berry K., Uterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., Deboy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidson I.,
RA White O., Fraser C., Collier A.;
RT "Complete sequence of *Pseudomonas syringae*."
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE016863; AA055747.1; --
DR TIGR: PSPTO2231;
DR InterPro: IPR006530; YD.
DR Pfam: PF05593; RHS_repeat; 6.
KW Complete proteome.
SQ SEQUENCE 982 AA; 106800 MW; 00E7CA6B93E19B78 CRC64;
Query Match 10.5%; Score 932.5; DB 16; Length 982;
Best Local Similarity 29.7%; Pred. No. 6.7e-42;
Matches 285; Conservative 142; Mismatches 357; Indels 175; Gaps 33;
QY 642 ANLTLYDELNNLQDNRPFVITTDVNGQLNRFDFGAGRVHVSQCLKSDGCKFYTI 701
Db 31 SSTLGYRD---DWNQR--CCITDDN-VQTYEYSDPIGSDVHK-----GPIQKT 74
QY 702 HTQOYDEGRHHTSYDYLTNGRQOTDPDKVHLSMSKSYDNNWQIANTHSYGVSEKIT 761
Db 75 WKQSGDPEGR-----ISGRSET-----WNLN-----FGKPDRI 103
QY 762 VDPIILTATKQLQSNNVQTKVTTTPSQOQITLFDGAGHLQSLCHTLTRDGDWRV 821
Db 104 ---TLTAGTGRSRTSHMSRSRLTT---TEQEL-----SRQFLYDGLGRC 144
QY 822 RKETDAIQCTIYQYDNNVRIQITLPGTIVNKYAPFSTDLITDIRV---NGIS--- 875
Db 145 TEQRDALQOQSTLFSYDNNRSMVSSSTLADGVSINRSYAPQSSSELATMLEVHQNTTIV 204
QY 876 LGQOQFDGLSRLTQSDQGRVWAYTSAQNDQCPSTVITPDGOFHYQYQPELDDAVLQV 935
Db 205 AGTQKFDGLERVTOIKTGDVREQYFNDAGEMQ--PSRTITAGLDNINFTYTRALTQIFSS 263
QY 936 ASNEITQOFSYNPVTKALLKAVAGQSUTPIYPSGRKMKENINDMKMSYLV-----T 989
Db 264 TAPDETAKFDYDKTSARLIEATNPQGTTRYDVHNLQTLGETWDL--LGQAWETRHOSS 321
QY 990 LRGLENGYDGL-----TGTIOKISRDTGRVTQIKDSSIKTLYDDLNRRHIGSQVDTLA 1044
Db 322 LLGRPIKRTDCLKGGAAGAEATRYDITLGRINFINQSLRTIITIDVLGOLCKVATEDLQ 381
QY 1045 TGHMLTTTVEFDGLNREIGRKLCDSSGHTLTIQOSWLKTQOLANRIVKLVGLQRTQOYS 1104
Db 382 AGTVIIDMEYDDQGEILRTQTASNAALTLTQWAVDGLLKLTRDLQOAGSPLLHETFS 441
QY 1105 YDSRNLNQKCDGAECPDKYGHSVITQNFYDIYGNITACHTTFADGTEDHATPKFAN 1164
Db 442 YDPRGLRTLNVYLSLPRDELQREMTQIFSFDELDNITLCQRTFTDGTSTRAAFKYGS 501
QY 1165 P-----TDPQCLTEVHTHPDMPDNIRLKYDKAGRVINITHGNNTENFTYDTLGRL--- 1216
Db 502 PGDDXKHQRCQLLSIATYTPPKTPDPTFSYDANGN--QLKDEHGN--SLHYDSQSRLQV 557
QY 1217 -QNGQG--SVGYDPLNRLVSOXTLDCEL--YYRETMVNEVNGEMIRLLRTGETII 1271
Db 558 AETGAPISQYRYDGHNLVATR-DGNESEILRFYEGHQLGSTVQEDQRTQYLHIGEOPL 616
QY 1272 AQQ---RASKVLLTGTDSQSVILTSQNLQSEAYSAVGHKSTANDASILGYNGERAD 1328
Db 617 GQOIVDDAEQTLILLITDANQSVMGFPQQGLEKAVYSAYGERHSEALLSIFAGFNGVRE 676
QY 1329 PVSQVTHLNGVRSYDPTLMRFHTPDLSPPFGAGGINPYSYCLGDFINRSDPSGH----- 1383

```

Db 677 AANGWYLLNGGYRAYNELLMEFHSDFLSPFAEGVNPYTYCLGNPIALRDPDTHGDASGQ 736
Qy 1384 -----LSWQ-----AWTIGMGIA-----GLLTATATGMAIAAAGI----- 1416
Db 737 TGRLLRDEGALPQOQGGDIMGVGVGVFVLGVAAITATLTATVTPVTVLGI 796
Qy 1417 -----AAAIASSTTTALAFG-----ALSVTSDITSIVSGALEDASPKASSILGWVS----- 1462
Db 797 SNTASAAAAGVSTGALIVGTALTAASTTANTVAIVNN-----DOTAGVEVGWGLGIAAV 851
Qy 1463 ---MGMAAGALAESAIKGGFKLA-----THLGAFAEDGENALLKSTSSSR 1505
Db 852 PVGLVGFAGAVARAAAKVAANAGTIGVRSVRIG-LAAAGARTISSAASAR 909

RESULT 9
ID Q87VG6 PRELIMINARY; PRT; 1669 AA.
AC Q87VG6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE YD repeat protein.
GN PSPT04970.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Bezy K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., Deboy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Hatt D., Selengut J., Nelson W., Davidsen T.,
RA White O., Fraser C., Collier A.,
RA "Complete sequence of Pseudomonas syringae."
RT Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE016874; AAC583398.1; -.
DR TIGR: PSPT04970.
DR InterPro: IPR005530; YD.
DR Pfam: PF05593; RHS_repeat; 5.
DR Complete proteome.
SQ SEQUENCE 1669 AA; 184268 MW; BC668B6C1E3FB7EE CRC64;

Query Match 10.4%; Score 920.5; DB 16; Length 1669;
Best Local Similarity 25.9%; Pred. No. 7e-41;
Matches 440; Conservative 212; Mismatches 662; Indels 387; Gaps 84;

Qy 152 AFDLKVPKLLNPAGHAIYIDMPEATQPELNRIYDDLGDHIDPLNLEYYQGL---IKTI 208
Db 72 AFGQLLPSEVTSPOGGTSLSEWRNQOQFLKKI--SSDGNVI--LSSTYSAPSSDSTIV 127
Qy 209 -LTLFPQKSGYRTELFRLNQLNSIHNFSLGNENPLTWSFGYTPIGKNGILGQWITSMT 267
Db 128 EISVWPKTDEAYTVRLSELSTALGTTRIEKNLKVSKLKGVC---ADPTLDRVLNRIE 184
Qy 268 APGLKETVNYNNQGHFPQSANLPVLPVYVTLKMQVPGAGQPAIQAEYSYTSNHYVGG 327
Db 185 EEDGSIELVYREG--GMPFFYRQ--PPLPCVTLHSLFPGACQSNITDHYFSGTNYLGF 241
Qy 328 GSGNGIWNKLDNLYGLMTEYNSGTSRRYKDEGH-----DQIVRIERTYNNYL 378
Db 242 SEPP--DAHONRLY-----YERLELRELVDDGYVQLRQNPDIHVSTRHAFNKNL 292
Qy 379 LTSECKQONGYIQTETAYVIAIGHNFDSPQSQ--FQLPK--TKTETWSADNSYRSEIT 434
Db 293 QVRE-----DLQVQFOAEKVISWEFANASPGKVFGLPKITTDYDLSHPNTERTTV 346
Qy 435 ETTFDESGNPLTKVIKDKTKQKISPSIHTWYPPAG-----EVD-----NC 476
Db 347 QTLAYNNIGQLTKSI-----AVDGVVTEWLYPYDPGGQGLDISLIAEKPLSKDLVTLTC 400

```

```

Qy 477 PPEPYGTFRFVKKITQITPYDSEFKDDP-----ERFIOYRY-----SLIGS 516
Db 401 PKVSEGY-----MPVKVEYVHDPAYPKSKQITAVAYQARENPNVQSVLPVSTVV 452
Qy 517 QSHVTLKIEERHYSATQLLNSTLFOYNTDKSELGELLKQTECTKG-----ENKTY--S 568
Db 453 LTGVTL-----DRTTWLPSLM-----EGKNALIEQRVISRIPDAVENTTAMKES 499
Qy 569 VVHKFTYTKQDDTLQOCSHSITTHDNFTIHR-----SOVRYRTGRLESDTDTK 616
Db 500 VVQNSWLWQQRNTL--TTSMLYDDNPFVGRVRAEAGQKIISRSFSLSGRPLSET-RD 556
Qy 617 DIVTQMSYDKLGRLLTR---TLNSGTFYANTL--TYDIEL--NNLODNNRPPFVITTDVN 670
Db 557 GLEFHYVHDSLGRIVRQERGTTTEAGAWKADAVETTDYSITAELQ-----VIVTEAE 608
Qy 671 GNQLNEFDGAGRHSVQCLKDS--GDGKFYTIHQOYDEQGRHHTSTYSVLYTNROOTD 729
Db 609 -QQVRLYDGLQRPWWVAIKTILPDSAFVCSIEYDGPDPATNQTLY--DYLPGELARTK 665
Qy 730 PKVHLSMSKSYDNGWQIANTHWSYG--VSEKITVDPIITLTKQLQOSNNNVOTGKEVT 787
Db 667 DARPEAAVDASKLAW--MADYTREDAGILNEQVICADSGAQLIRQLSGRLNDSHTALLE 725
Qy 788 TYTFS-----QQPIQITLFDFAHLSQSCHTLTRDGDWRVRKETDALGCTIYQYDNYN 840
Db 726 TLRPSTARDASTDRTIERT--FDERRLIKRT-----SNTSEHC--JEYDELE 770
Qy 841 RVIQITLPGDTIVNRKXAPFSTDLITDIRVNGISLQOQTFDGLSRLTSQDQGRVWAYT 900
Db 771 RAVALLIADPTGRTERKXHOLS--DVIQLNVGVSVLGTQKMTAAARQTTVGE-----LTYE 824
Qy 901 YSAGNDQCFSTVITPDGOFIHYQVQPELDDAVLQVNASNEITQQ-----PSYNNVT- 950
Db 825 PFGGS---ASTVVRPDKTLLSASVADGHTATLSINKKVHTQVVISQPNVLTVTVDPSV 881
Qy 951 ----GALLKAVAGQSILTPYY--PSGRKLVENINDMKMSYLWTLRGLENGVTDTLGT 1003
Db 882 PSAEAWSSLTSSPOSGLGTSITQTSPRGSRQAEWTRSLKG-----RLLTN--TAVDGR 932
Qy 1004 IQKISRDTHGRTVQIKOSSIKITLYDLDNRHIGS-----QVTDLATGHMLTTVEFPLGN 1059
Db 933 QMRVFRDYLDVRVRV----ILGELHYLWLSAFGEPLQRTVVNOASGERLDVRFETWDAFG 988
Qy 1060 REIGRKLCDSGHTLIDIQOSWLKTTQOLANRIVKLNGVLQRTQOYSYDSRNLNAYKC--D 1117
Db 989 QEIAREYTLNNKPLLALNTNSVLANGOVSSKTLTREGVLQRTGEGFSYDARDLNSYECTD 1048
Qy 1118 GAEPTDKYGHISIVTQFTYDIYGNITACHTTFADGTEDHATPKF---ANPTDPCQL--TE 1173
Db 1049 VADWPQDQAGKSLQSGYGYDELHNLSECSSTYADGSTCIQTYTYDTYKXPTFRLSVKTE 1108
Qy 1174 VHH--THPMDPNIRLKYDKACRVINITDNGNTENFTYDTLGRL---ONGQGSV---YGY 1226
Db 1109 LRSGSTTSTOTATLADANGN--OTTDESQRT--LAYTPLGLASVKDNDKLLITYSY 1164
Qy 1227 DPLNRLVSQKDTTL--DCELYRETMLVNEV-----RNGEMIRLLRT-----GETII 1271
Db 1165 DAFGRLLISQYIGATKHTCELLYDGTQLTGEAWFDANREFKRILFSDVMVQTCIGETV- 1223
Qy 1272 AQQRASKVLLTGTDSQOSVILLTSKQ-----NLSQEAYSAGKHKSTANDASI-LGVNGE 1325
Db 1224 ---RSDFVL---TDPGGVGVFSADGTAGVKLHPLGTYPG--ESTNLSGGRLGNSE 1275
Qy 1326 RADPVSGVTHLNGYRSYDPTLMRFHTDPSLSPFAGGINPYCYCLGDPINRSDPPGHL- 1384
Db 1276 RIDPVLGWHLNGYRTYSPAQRHLQPDSSWSPFAGGINNTAYCAGDPVNLFPDPSGHVM 1335
Qy 1385 --SQA-----WTGI-----CMGIAGLLLTATCGMAIARAAGGI 1416
Db 1336 ISRWGASNMISDLTKALQETSPOQLHFHWRGLAVNASVAVAGVLMVPLTGTGSSLGPAAGV 1395

```

```
QY 1417 AAATASTTALAFGALSVTSDITSIVSGALEDASPK-----ASSILGVSVGMGAAG 1469
DB 1396 LA-----TTLAVASAGL-----DIASYV---LEDVNPRLARKLGTAAALGFISNAPPRAG 1443
QY 1470 L-----AESAIKGGTKLATHLCAF--AEDGENALLKSTSESSRIK 1507
DB 1444 LRLGRLRLRWTSSTGRLISREVESIFTKGPKLLTGTIKVLKFG-----MSSVGSASNVTR 1499
QY 1508 WGVTRSLDREIVRNEEGQVINKDHSRGYTDNFMKGEOAILVHGDKG-----FLYHTEGKN 1563
DB 1500 W-----APGKLDKF-EKLENALFMTVDKQKRVTFMAH--GVK 1534
QY 1564 -----HNGKGPYT--RHTPEQLVDVLKDNINIVDLTQGGDKPVHLLSCVYGS 1607
DB 1535 PDDVGDAMMAIBSYSGSPVSGKSFAYSFFDFELKSKK--VDLNNK--YEKVRLLIMCHSAD 1591
QY 1608 SGA---ADKVKAKYINRPVIAY 1625
DB 1592 GGEKSFATFSKLTNPKVKGY 1612

RESULT 10
ID Q45905 PRELIMINARY; PRT; 505 AA.
AC Q45905;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Orf 505 protein.
GN ORF 505.
OS Coccidia burnetii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coxiellaceae; Coxiella.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Scurry Q217;
RX MEDLINE=97294470; PubMed=9150226;
RA Willems H., Ritter M., Jaeger C., Thiele D.;
RT "Plasmid-homologous sequences in the chromosome of plasmidless
ET Coccidia burnetii scurry Q217.";
RL J. Bacteriol. 179:3293-3297(1997).
DR EMBL; X93204; CAA63682.1; -.
SQ SEQUENCE 505 AA; 57841 MW; 8B60B4C307CAB8D0B CRC64;

Query Match 10.4%; Score 919.5; DB 2; Length 505;
Best Local Similarity 41.1%; Pred. No. 1.2e-41;
Matches 217; Conservative 75; Mismatches 195; Indels 41; Gaps 16;

QY 18 NEF-FTQANNFTSAVGGVDPRGTGLYNIQITLGHVGNLGNLPTLPLTLYSPLNKTDI- 75
DB 2 NELPYQTATNFISAVQGGVDPRGTGLFTVMVLAELTGNLGNLGPDLFTLNYSPILSTSNIC 61

QY 76 GFGIGFNLGLSVYDRKNSLLSLSTGENYKVIETDKTVKLOQKLDNLRPEKDKENCYRI 135
DB 62 GFGIGCSVGSISYDKNNKLLILSSERYKTEDWNSVYVQKINNFKFEK--IKNGYII 119

QY 136 IHKSGDIEVLTFGNNAFDLKVPKLLNPAGHAIYIDWNFEATQPLRNRIYDDLGDHDP 195
DB 120 KYKNGKTEYLYKYGDNLF---LPQKIFSLGWPLKLTWENRQYVNLTKIEDAKD---V 172

QY 196 LLNLEYQGLIKTILTLFPQKEGYRTELRLNRLNLSIHNFSLGNENPLTWSFGYTPIGK 255
DB 173 LCKIDYQFSDWARIITFWPGKTESYTFQLDFVNELYLWYTNKTSRE--LWMSFNDDVGA 230

QY 256 NGILQWITMTAPGGLKETVNYNNQGHFFQSANLPVLVYVTLMKVFGAGQPAIOA 315
DB 231 GNFT---LTQVKSPTGLTETVNYQAGVM--RFPDESCKPALPSVYNYRQSPGMQGPDIK 285

QY 316 EYSTSHNYVGGGNS--GIWNNKLNLYG-LMTEYNYGSTESRRYKDKEGHDQIVRIERT 372
DB 286 EYEYASNYLGGASLGAKWNEDENIYNVNMDDITYSSTEKLIVDNRE---LVSISRI 341

QY 373 YNNYHLLTSECKQNGYIQTETAYYALIGHNFDSPSQFQLPKTKTETWR--SADNSYRS 431
DB 342 YNSYLLISETTRQNSCEVIVETDYVAKPGLSFDQPKQFOLPKKEKKTWRENSKNQCRS 401

QY 432 EITETTFDESNGPLTKVIKDKTKQIKIISPSTHWEYYPAGEVDN---CPPEPYGTFRV 487
DB 402 EITTTTFDEPNLLTKIEPD-----GKTETIYYDSKGETDKGIVLCPPEPNGFVRPV 454
```

```
QY 374 NNYHLLTSECKQNGYIQTETAYYALIGHNFDSPSQFQLPKTKTETWR--SADNSYRS 432
DB 342 NSYLLISETTRQNSCEVIVETDYVAKPGLSFDQPKQFOLPKKEKKTWRENSKNQCRS 401
QY 433 ITETTFDESNGPLTKVIKDKTKQIKIISPSTHWEYYPAGEVDN---CPPEPYGTFRVK 488
DB 402 ITTTTFDEPNLLTKIEPD-----GKTETIYYDSKGETDKGIVLCPPEPNGFVRPV 454
QY 489 KTIQTPYSEPKDDPEKFIQYRYS-----LIGSQSHVTLKIERHYSAT 532
DB 455 TOIVTPANSEFY-APVQOTTAYAOYPCITAGSSLSYAVLQNKHFAVT 501

RESULT 11
ID Q9X626 PRELIMINARY; PRT; 526 AA.
AC Q9X626;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Coccidia burnetii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coxiellaceae; Coxiella.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1140;
RA Radomski K.U., Willems H., Lautenschlaeger S., Jaeger C., Baljer G.;
RT "Sequence of OpDV plasmid.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF131076; AAD33498.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 526 AA; 60111 MW; AF26E1D9E1B56834 CRC64;

Query Match 10.4%; Score 919.5; DB 2; Length 526;
Best Local Similarity 40.8%; Pred. No. 1.3e-41;
Matches 221; Conservative 73; Mismatches 203; Indels 45; Gaps 16;

QY 18 NEF-FTQANNFTSAVGGVDPRGTGLYNIQITLGHVGNLGNLPTLPLTLYSPLNKTDI- 75
DB 2 NELPYQTATNFISAVQGGVDPRGTGLFTVMVLAELTGNLGNLGPDLFTLNYSPILSTSNIC 61

QY 76 GFGIGFNLGLSVYDRKNSLLSLSTGENYKVIETDKTVKLOQKLDNLRPEKDKENCYRI 135
DB 62 GFGIGCSVGSISYDKNNKLLILSSERYKTEDWNSVYVQKINNFKFEK--IKNGYII 119

QY 136 IHKSGDIEVLTFGNNAFDLKVPKLLNPAGHAIYIDWNFEATQPLRNRIYDDLGDHDP 195
DB 120 KYKNGKTEYLYKYGDNLF---LPQKIFSLGWPLKLTWENRQYVNLTKIEDAKD---V 172

QY 196 LLNLEYQGLIKTILTLFPQKEGYRTELRLNRLNLSIHNFSLGNENPLTWSFGYTPIGK 255
DB 173 LCKIDYQFSDWARIITFWPGKTESYTFQLDFVNELYLWYTNKTSRE--LWMSFNDDVGA 230

QY 256 NGILQWITMTAPGGLKETVNYNNQGHFFQSANLPVLVYVTLMKVFGAGQPAIOA 315
DB 231 GNFT---LTQVKSPTGLTETVNYQAGVM--RFPDESCKPALPSVYNYRQSPGMQGPDIK 285

QY 316 EYSTSHNYVGGGNS--GIWNNKLNLYG-LMTEYNYGSTESRRYKDKEGHDQIVRIERT 372
DB 286 EYEYASNYLGGASLGAKWNEDENIYNVNMDDITYSSTEKLIVDNRE---LVSISRI 341

QY 373 YNNYHLLTSECKQNGYIQTETAYYALIGHNFDSPSQFQLPKTKTETWR--SADNSYRS 431
DB 342 YNSYLLISETTRQNSCEVIVETDYVAKPGLSFDQPKQFOLPKKEKKTWRENSKNQCRS 401

QY 432 EITETTFDESNGPLTKVIKDKTKQIKIISPSTHWEYYPAGEVDN---CPPEPYGTFRV 487
DB 402 EITTTTFDEPNLLTKIEPD-----GKTETIYYDSKGETDKGIVLCPPEPNGFVRPV 454
```

```
QY 488 KKIQTPTVDSF-----KDDPEKFIQYRSLIGSQSHVTLKIBERHYSATOLLNSTLFPQYN 543
Db 455 KTIQIVTPADSEFYAPVQQTYYAYAQYPCIAAGSSLSYAVLQTQETLCSDDVL-----LLTIN 510

QY 544 TD 545
Db 511 TD 512

RESULT 12
Q52883 PRELIMINARY; PRT; 526 AA.
AC 052883;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
GN ORF 526.
OS Coxiella burnetii.
OG Plasmid QpRS.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coxiellaceae; Coxiella.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Priscilla 0177;
RA Lautenschlaeger S., Jaeger C., Willems H., Baljer G.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y15898; CAA75844.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 526 AA; 60104 MW; 42D04FF8CC361CC CRC64;

Query Match 10.3%; Score 916.5; DB 2; Length 526;
Best Local Similarity 40.8%; Pred. No. 1.8e-41;
Matches 221; Conservative 73; Mismatches 203; Indels 45; Gaps 16;

QY 18 NEF-FTQANNFTSAVSGVDPRDTGLYNIQTILGHVGNGLPGLTPLTSLYSPLNKTDI- 75
Db 2 NELPTQATNFISAVQGVDPRTGLFTVNMVLAELTGNDNLGDPDLFTLNSPLSTSNIC 61

QY 76 GFGIGFNLGSLVDRKNSLSLSTGENYKVIETDKTVKLOOKLDNLRFKDLKENCYRI 135
Db 62 GFGIGCSGIGSIYDKNNKLLLSGGERYKIEDMDNGVYVRQKINNFKFEK--IKNGYII 119

QY 136 IHKSGDIEVLGTFNNAFDLKVPKLLNPAGHAIYIDWNFEATQPLNRIYDLDGHDIP 195
Db 120 KYKNGKTEYLYKYGDNLF---LPQKIFSTLWPLKLTWENRGQVYVNLKIEDAKD---V 172

QY 196 LLNLEYQGLIKTILTLFPQKEGYRTELRFNLNQLNSIHNFSLGNENPLTWSFGYTPIGK 255
Db 173 LCKIDYQFSDWARITFWPGKTESYTFQDLDFVNELYLWVTKNKSRE--LWVSFNDDVGA 230

QY 256 NGILGQWITSTAPGGLKETVYNNNGHFFQSANLPVLYVTLMKQVPGAGQAIQA 315
Db 231 GNFT---LTQVKSFTGLTETVNYQAGV--RFPDESGKPALPSVYNYRQSPGQPDIVK 285

QY 316 EYSTSHNYVGGGN--GIWNNKLDNLG-LMTEYNGSTESRYKDKGHDQIVRIERT 372
Db 286 EYEYASNYLGYGASLGKAWNEDEDNINVMDDYTSSTEKLIVDNRE---LVSIISRI 341

QY 373 YNNHLLTSECKQNGYIQTETAYAIIGHNFDSPQSPQLPKTKTETWR--SADNSYRS 431
Db 342 YNSVYLLISETRQNSCEVIVETDYAKPGLSPDKQKQQLPKKEKKTWRENSKNQCRS 401

QY 432 EITTFDESQNPITKVIKDKTKIISPSHWEYYPAGEVDN-----CPPEYFGTRFV 487
Db 402 EITTTTFDEPNLLTKIEPD-----GKTEYIYDYSKGETDKGIVLCPEPENGFRV 454

QY 488 KKIQTPTVDSF-----KDDPEKFIQYRSLIGSQSHVTLKIBERHYSATOLLNSTLFPQYN 543
Db 455 KTIQIVTPADSEFYAPVQQTYYAYAQYPCIAAGSSLSYAVLQTQETLCSDDVL-----LLTIN 510
```

```
QY 544 TD 545
Db 511 TD 512

RESULT 13
Q45949 PRELIMINARY; PRT; 528 AA.
AC Q45949;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Orf 528.
OS Coxiella burnetii.
OG Plasmid QpH1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coxiellaceae; Coxiella.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nine Mile phase I;
RA Thiele D., Willems H., Haas M., Krauss H.;
RT 'Analysis of the entire nucleotide sequence of the cryptic plasmid QpH1 from Coxiella burnetii.';
RL Eur. J. Epidemiol. 10:413-420(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Nine Mile phase I;
RA Thiele D.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; X75356; CAA53130.1; -.
DR PIR; S38242; S38242.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
KW Plasmid.
SQ SEQUENCE 528 AA; 60422 MW; 492589F2DAA890A6 CRC64;

Query Match 10.3%; Score 912.5; DB 2; Length 528;
Best Local Similarity 40.8%; Pred. No. 3e-41;
Matches 222; Conservative 72; Mismatches 203; Indels 47; Gaps 17;

QY 18 NEF-FTQANNFTSAVSGVDPRDTGLYNIQTILGHVGNGLPGLTPLTSLYSPLNKTDI- 75
Db 2 NELPTQATNFISAVQGVDPRTGLFTVNMVLAELTGNDNLGDPDLFTLNSPLSTSNIC 61

QY 76 GFGIGFNLGSLVDRKNSLSLSTGENYKVIETDKTVKLOOKLDNLRFKDLKENCYRI 135
Db 62 GFGIGCSGIGSIYDKNNKLLLSGGERYKIEDMDNSVYVRQKINNFKFEK--IKNGYII 119

QY 136 IHKSGDIEVLGTFNNAFDLKVPKLLNPAGHAIYIDWNFEATQPLNRIYDLDGHD 193
Db 120 KYKNGKTEYLYKYGDNLF---LPQKIFSLGFWPLKLTWENRGQVYVNLKIEDAKD--- 173

QY 194 IPLLNLLEYQGLIKTILTLFPQKEGYRTELRFNLNQLNSIHNFSLGNENPLTWSFGYTP 253
Db 174 -VLCKIDYQFSDWARITFWPGKTESYTFQDLDFVNELYLWVTKNKSRE--LWVSFNDDV 230

QY 254 GKNGILGQWITSTAPGGLKETVYNNNGHFFQSANLPVLYVTLMKQVPGAGQAI 313
Db 231 GAGNT---LTQVKSFTGLTETVNYQAGV--RFPDESGKPALPSVYNYRQSPGQPD 285

QY 314 QAEYSYTSNHYVGGGN--GIWNNKLDNLG-LMTEYNGSTESRYKDKGHDQIVRIE 370
Db 286 VKEYEYTSVNYLGYGASLGKAWNEDEDNINVMDDYTSSTEKLIVDNRE---LVSIIS 341

QY 371 RYNNHLLTSECKQNGYIQTETAYAIIGHNFDSPQSPQLPKTKTETWR--SADNSY 429
Db 342 RYNSVYLLISETRQNSCEVIVETDYAKPGLSPDKQKQQLPKKEKKTWRENSKNQ 401

QY 430 RSEITTFDESQNPITKVIKDKTKIISPSHWEYYPAGEVDN-----CPPEYFGTR 485
Db 402 RSEITTTTFDEPNLLTKIEPD-----GKTEYIYDYSKGETDKGIVLCPEPENGFRV 454

QY 486 FVKKIIQTPTVDSF-----KDDPEKFIQYRSLIGSQSHVTLKIBERHYSATOLLNSTLFPQ 541
```



```
|||||
455 FVKQIVTADSEFAPVQQTYYAQAQPCIAAGSSLSYAVLQTOETLCSDVLT-----LIT 510
QY 542 YNTD 545
Db 511 INTD 514

RESULT 14
Q45902 PRELIMINARY; PRT; 589 AA.
AC Q45902;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Orf 589 protein.
GN ORF 589
OS Coxiella burnetii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coxiellaceae; Coxiella.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Scury Q217;
RX MEDLINE=97294470; PubMed=9150226;
RA Williams H., Ritter M., Jager C., Thiele D.;
RT "Plasmid-homologous sequences in the chromosome of plasmidless
RL Coxiella burnetii scury Q217."
RL J. Bacteriol. 179:3293-3297(1997).
DR EMBL; X93204; CAA63679.1; -.
DR InterPro; IPR006530; YD.
DR Pfam; PF05593; Rhs repeat; 2.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 3.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 3.
SQ SEQUENCE 589 AA; 64241 MW; 8B6BEE8BEDE9A035 CRC64;

Query Match 10.0%; Score 884; DB 2; Length 589;
Best Local Similarity 38.0%; Pred. No. 1.3e-39;
Matches 232; Conservative 85; Mismatches 182; Indels 112; Gaps 24;

QY 1103 YSDSRNLNQYKCDGACETDKYGSIVTQNTFYDIYGNITACHTTFADGTEDHATFX 1162
Db 2 FAYDSRNLLEYTCNGEARPQDPYGAHRTQTSYDALGNMTKTQTDPSGG-RNTATYII 60

QY 1163 ANPTDPCQLEVVHHTHMDPNDRLKYDKAGRVNITDNGHNTENFYDTLGLQ--NGQ 1220
Db 61 -SAIDPQLKVNNDHNYKEITLVDYDKAGMR--DEAGRT--LRYDALGLQVNGE 115

QY 1221 GS---VYGYDPLNRLVSQKT-DTLDCELYRETMLVNEVR--NGEMIRLLRTGETIIAQ- 1273
Db 116 GAKGGQAYDALNTLVQVWQDEPIYDLYRADDLVGEARRDSSQTRYVKSQCCVGC 175

QY 1274 --ORASKV-LITGTDSCQSVLITSDKONLSCEA--YSAYGKHSTANDASILYNGERADP 1329
Db 176 TKQSSNTSWLTTNQGSVLVSSEGNHAFQDCIYTPGYRTPTQETPSVLGNGERLDP 235

QY 1330 VSGVTHLNGYRSYDPTLMRFHPDLSLSPFGAGGINPYSCLDGPPINRSDPSGHLGWQAW 1389
Db 236 VSSIVHLNGYRAYNPILMRNCPDSSPFGAGGINPYAYCDGDPINRVDPNHLSQAE 295

QY 1390 TGIGMGAGLLLTATCGMAIAGGTAAGTAIASTTTALAFGALSYSDDTTSIVSGALED 1449
Db 296 LGIGLVGVGLVAVFTAGTSIAAGAAISAAIESASISLVVGLTGVAAADVASTAGALE 355

QY 1450 ASKRASSTLGVSMGMGAAGLAE---SAIKGCTKLATHLGAFAEDG-----ENALL 1497
Db 356 ANPOASATLWISLGLGPGAVSLATAARAGKL---ISLAKGGKIRSQSPVQGINY 412

QY 1498 KSTSESSIRKQVMT--RSLDREIVRNEE-----GOVTKHSRGYTNF----- 1538
Db 413 RSLRGDPLRGPPHFPQSLSRVTVAPESMRPAGLNHYKVSQKSLGYQHVFGADREIFG 472

QY 1539 -----MGKEQAII-----VHGDKGFLVHTEG-----NK 1563

Db 473 YEIREPIEFFRRRSITKRDIVILSGTHGRVHGDN-----WTSQGLRPDILERAFLYEDV 528
QY 1564 HNGXGPYTRHTPEQLVDYLDKNNIVDTQGG-----DKPVHLLSCYKSGSAAKMA 1615
Db 529 QNYXG-----QLNGRVKIVDMAGMTESEFGHVRNVNQHVLGYCWRNDEA---LT 577

QY 1616 KYIN-RPVIAV 1625
Db 578 YHRLRPIISY 588

RESULT 15
Q931J8 PRELIMINARY; PRT; 762 AA.
AC Q931J8;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Vibrio sp. CH-291.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=161725;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CH-291;
RA Jones J., Lewin A., Appel B.;
RT "Cloning of a hemolysin encoding region of a Vibrio species.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ314791; CAC40978.1; -.
DR InterPro; IPR006530; YD.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 2.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 2.
KW Hypothetical protein.
FT NON_TER 762
SQ SEQUENCE 762 AA; 87823 MW; A172ABBCD447D2CA CRC64;

Query Match 7.3%; Score 648; DB 2; Length 762;
Best Local Similarity 26.1%; Pred. No. 1.2e-26;
Matches 222; Conservative 150; Mismatches 341; Indels 136; Gaps 32;

QY 22 TOANFTSAVSGVDPRTGLYNIQITLGHIVGNLGLPTLPLTLSYSPLNKTDIGFGIGF 81
Db 6 SNAFNFSFVSGVDPRTGSYSISFTLGLLSNKLKSGNFKLITISHYLNKVDGEGFLGW 65

QY 82 NPLGSVYDRKNSLLSLSTGENYKVI---ETDKTVKLOOKLNDLRF-----EKDLKENCYR 134
Db 66 SISMSYDYKISRKLSLSSGRTPETVLKSDSNELIILHRKTKDRAFLVENEREIK- 120

QY 135 ITHKSGDIEVLTFGNNAFDLKVPK--KLLNAPAGHAIYIDNWFATQPLRNKRYDLDLGH 192
Db 121 VVYIDGKVEYI-----DYESGKLIETVSNLGHIEIFYRYFNGLLSLSKISQDF-GH 171

QY 193 DIPL--LNLEYQGLIKTILT-----LFPQCKEGYRTELRLNRLQNSIHNFLSGNEN 242
Db 172 SITIDHNSKYVTIINSVAENDIYKRNVLKLSHGY-----GRILTSI---SWANSS 221

QY 243 PLTWFGYPTPIKNGILGOW-ITSWTAPGGLKETVYNNNOCHHFP-OSANLPLVLPYVT 300
Db 222 NLKTTIDYKIDK---LGNYAIQVKNYSGLVETIEYS--YEGHLLPKKNKNFNPIPNVK 276

QY 301 LMKQVPGAGCPAIQAEYSYTSNHYVGGSGNINWKNLNDLYGLMTEYNYGSPESRRYKDK 360
Db 277 RHITLPGSQPKTVTEYFYSLSKNLYGYSNLLWEGDVLFTASDYRYSQEVW-----N 332

QY 361 EGHQDQIVRIERTYNNYHLLTSECKQONGYI-OTTATYVAIIGHNFDQPSQFQPLPKYT 419
Db 333 GSH-----IIRVYVKNYHLLSEKFKFNGVLYKEIDLEYVADLTQGIDEQPNYNYKXKES 388

QY 420 ETWRSDNYSRSEITETTFDESGNPLTKVTKDKTKQKLIISPSTHWEYYPAGEVDCNCPDE 479
Db 389 ITHF-LDGTQRTSEVSEYGFDDYGNCFECDDKQTT-----SPFEYPCGEQEQCPAH 440

QY 480 PYGFTFVKKIIQTPTDSEFKDDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQLNSTL 539
```


[illegible]

Search completed: July 3, 2004, 05:50:11
Job time : 77 secs